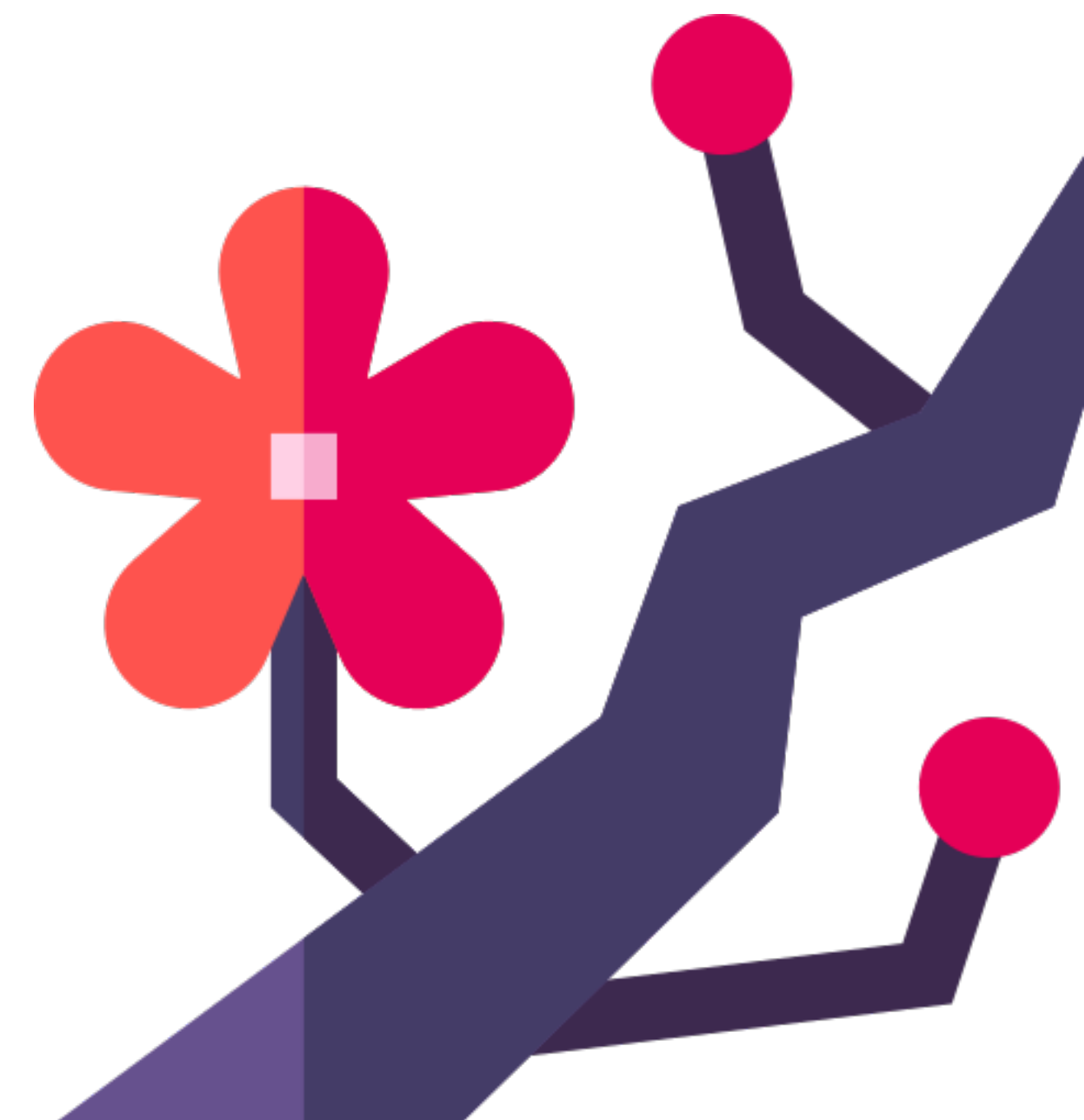


Etalab

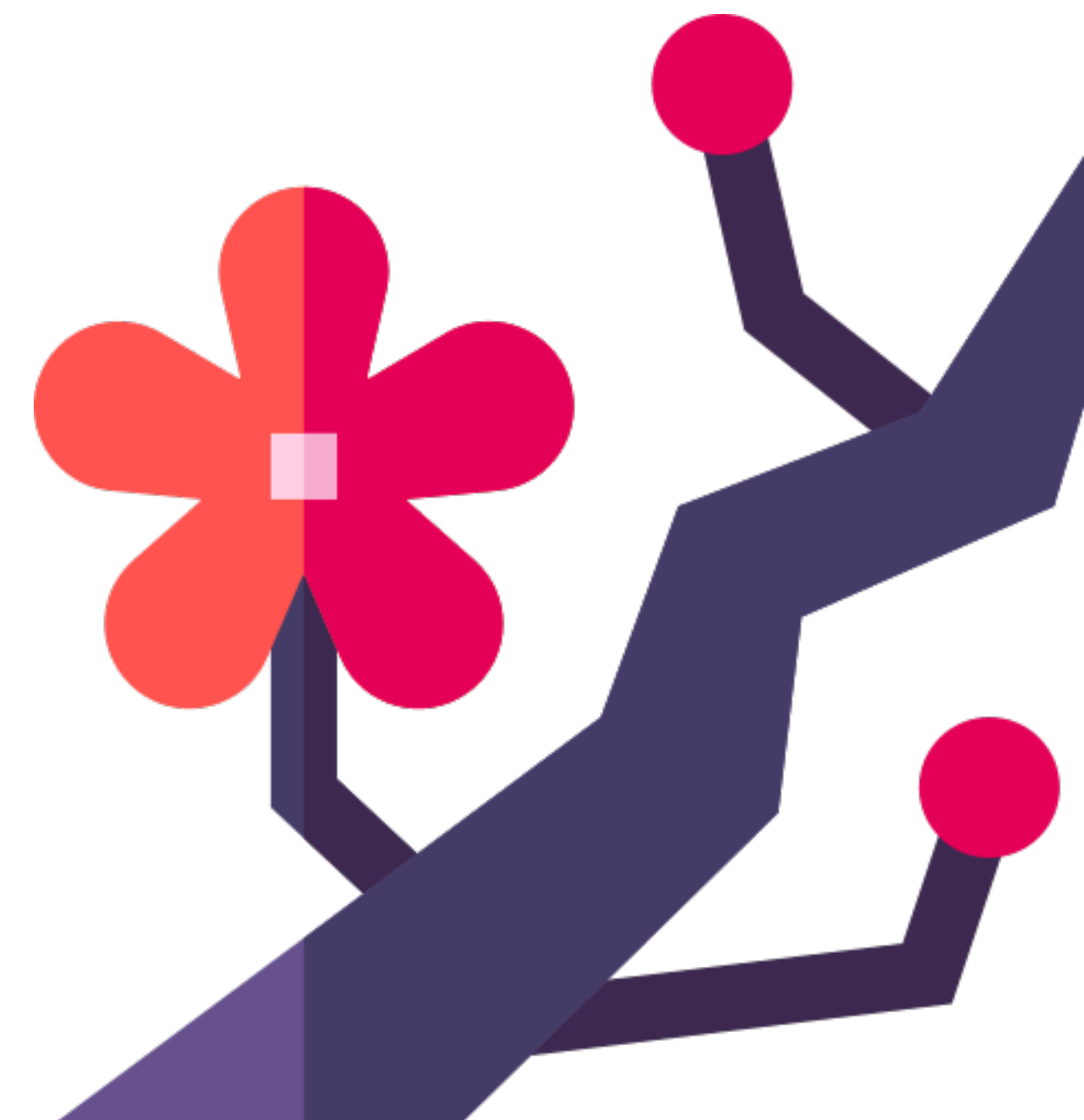
Cycle de vie de la donnée ouverte de qualité

Geoffrey Aldebert

Lundi 5 juillet 2021



Contexte



La mission



- **Plateformes ouvertes des données publiques**

Rassembler, ouvrir et partager les données publiques et les rendre accessibles à tous

- **Entrepreneurs d'Intérêt Général**

Ouvrir l'administration à des talents du numérique pour résoudre des défis publics en 10 mois

- **Sciences des données & Intelligence Artificielle**

Accompagner les administrations dans l'usage des datasciences et de l'Intelligence Artificielle

data.gouv.fr

37 000
Jeux de données

2 800
Organisations

75 000
Utilisateurs



data.gouv.fr

Plateforme ouverte des données publiques françaises

Données Réutilisations Organisations Tableau de bord Documentation Actualités Support

Geoffrey Aldebert



Recherche

Agriculture et Alimentation

Culture, Communications

Comptes, Économie et Emploi

Éducation, Recherche, Formation

International, Europe

Environnement, Énergie, Logement

Santé et Social

Société, Droit, Institutions

Territoires, Transports, Tourisme

Partagez, améliorez et réutilisez les
données publiques



CONTRIBUEZ !

200 000
Ressources

2 650
Réutilisations

8 800
Discussions

THÉMATIQUES À LA UNE

Données de

RÉFÉRENCE

Données relatives à la

COVID-19

Données relatives aux

COMPTES PUBLICS

JEUX DE DONNÉES



Données de laboratoires pour le
dépistage : Indicateurs sur les
variants (SI-DEP)

Information du 20 mai 2021 A compter de
ce jour, les indicateurs seront corrigés afin



Synthèse des indicateurs de suivi
de l'épidémie COVID-19

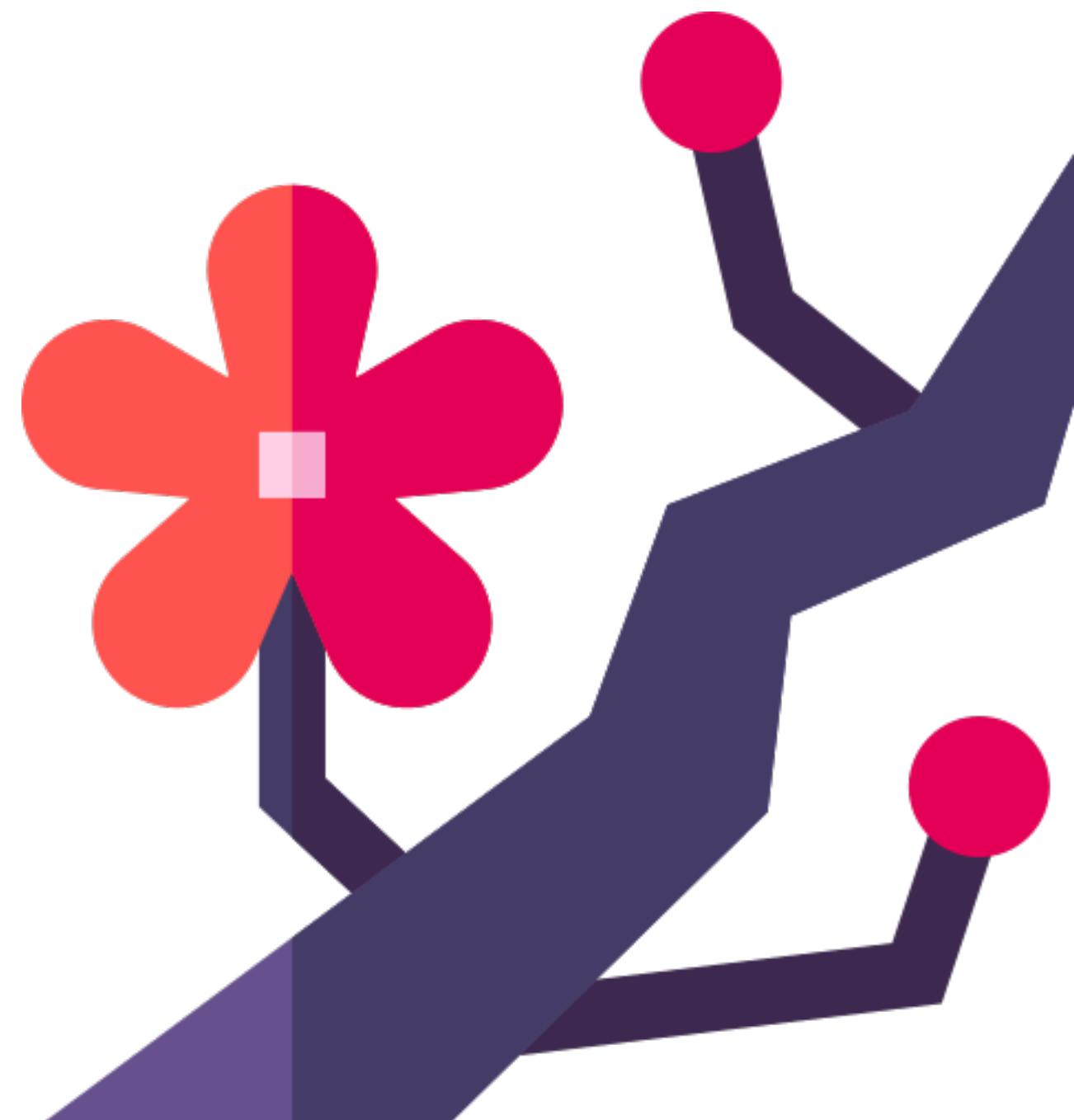
Description du jeu de données Ce jeu de
données comprends l'essentiel des
indicateurs de synthèse permettant le suivi



Données relatives aux personnes
vaccinées contre la Covid-19 (VAC-
SI)

MAJ 30 Avril 2021 L'estimation des
couvertures vaccinales pour les résidents

Qualité des données



Qualité des données

Les dimensions de la qualité

Complétude



Unicité



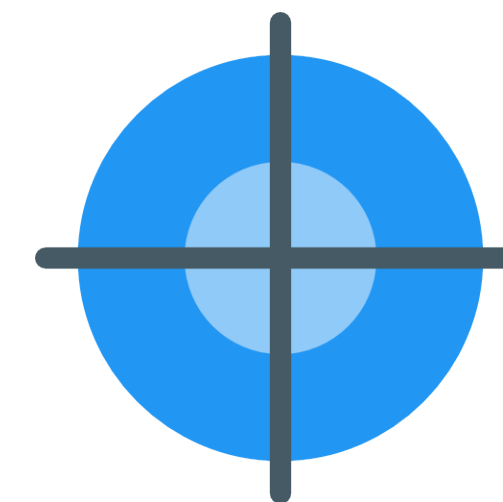
Consistance



Validité



Précision



Les dimensions de la qualité

Nombre de financement	Montant total	Année
5	1000	2010
8	1200	2011
7		2012
7	800	2013
6	800	2013
9	900	2014
9	8000	2015
8	1100	2016
8	-2000	2017
6	700	2018
139	1000	2019
6	1100	2020

Qualité des données

Les dimensions de la qualité

Nombre de financement	Montant total	Année
5	1000	2010
8	1200	2011
7		2012
7	800	2013
6	800	2013
9	900	2014
9	8000	2015
8	1100	2016
8	-2000	2017
6	700	2018
139	1000	2019
6	1100	2020



Complétude



Unicité



Consistance



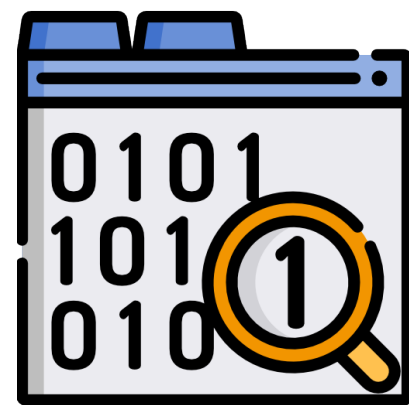
Validité



Précision

Qualité des données

Quels axes prioriser pour la qualité des données ouvertes ?



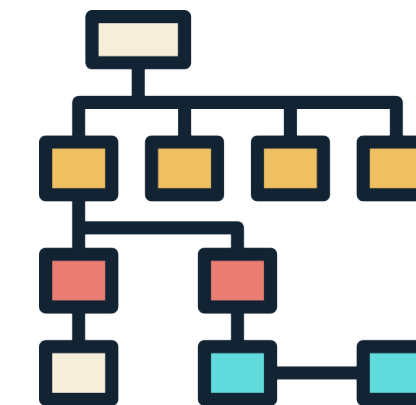
Contenu

Données identifiables clairement
avec un type de donnée simple



Format de fichier

Adaptés pour les réutilisateurs
de l'Open Data



Structure des fichiers

Noms des propriétés explicites,
compréhensibles et interprétables

Est-ce suffisant ?

Non, si nous sommes plusieurs à produire ce même type de données

Exemple fictif pour des données de type « espace vert »

Superficie (en m2)
Nombre d'arbres
Adresse
...



Surface totale (en ha)
Total plantes
Coord. géo.
...

Qualité des données



schema.data.gouv.fr

Aller sur data.gouv.fr

schema.data.gouv.fr est l'initiative de data.gouv.fr de référencement des schémas de données publiques pour la France.

Documentation

Schémas en investigation et en cours de construction

Guide : créer des schémas de données

Ajouter un schéma

Validation des schémas

Intégration avec d'autres systèmes

Schémas réglementaires

Schémas

Données essentielles des marchés publics français

Schéma des attributs des arbres urbains

DAE

Schéma d'aménagements cyclables

Spécification du fichier de déclaration de profil d'acheteur

Schéma IRVE

Qu'est-ce qu'un schéma ?

Les schémas de données permettent de décrire des modèles de données : quels sont les différents champs, comment sont représentées les données, quelles sont les valeurs possibles etc.

Cas d'usages des schémas

Les schémas permettent entre autres de valider qu'un jeu de données se conforme à un schéma, de générer de la documentation automatiquement, de générer des jeux de données d'exemple ou de proposer des formulaires de saisie standardisés.

Vers des données de qualité

Véritables gabarits, ils permettent une montée en qualité des données proposées en open data, surtout lorsque plusieurs producteurs de données produisent un même jeu de données.

Une plateforme de référencement nationale permet un accès aux schémas produits par différents acteurs et facilite l'intégration avec des systèmes informatiques par le biais de standards, d'URLs stables, de processus de validation et d'API.

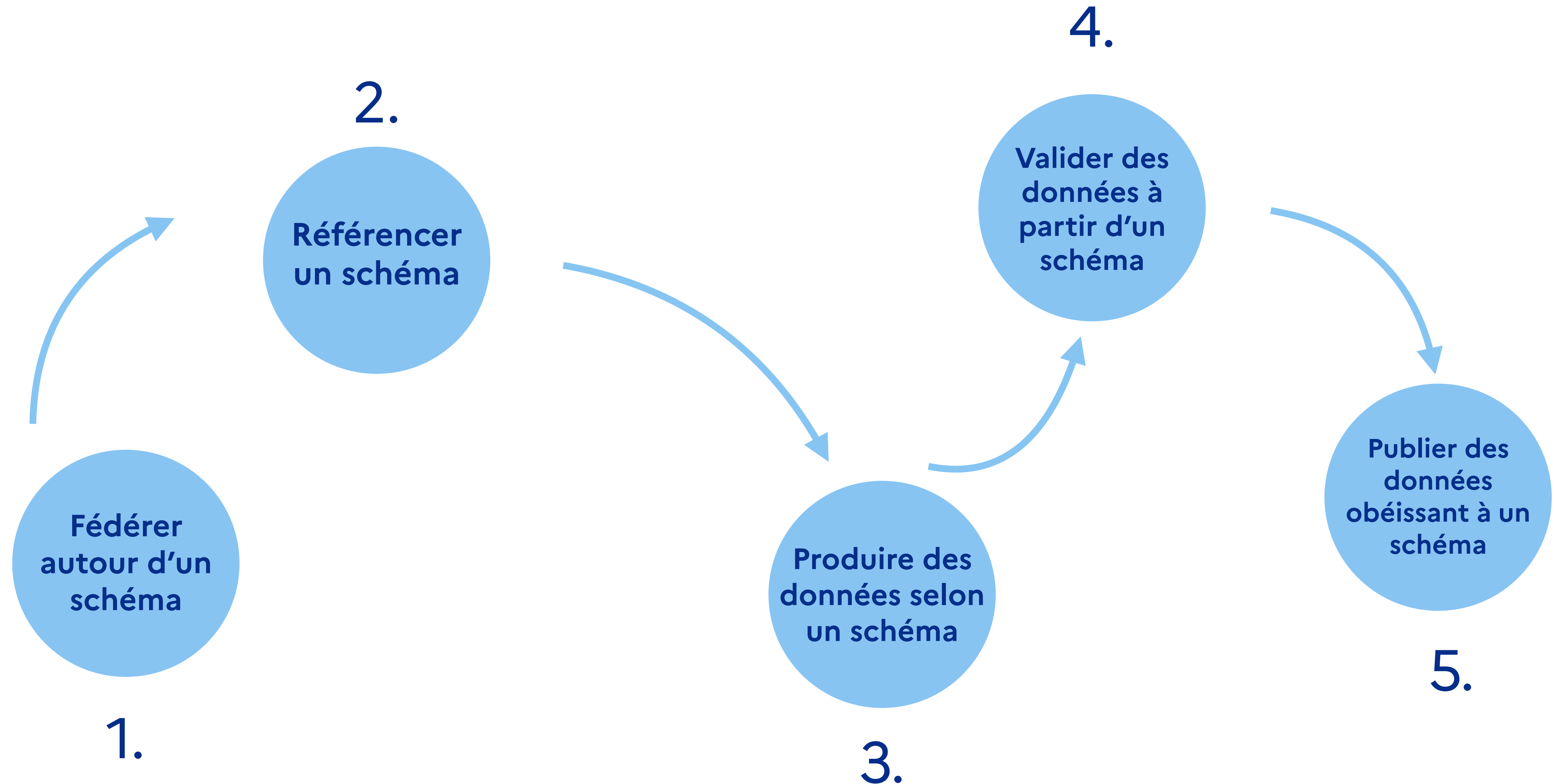
Nous vous invitons à [prendre connaissance d'un schéma](#) ou découvrir quelle est la procédure pour [ajouter un schéma](#).

Lancée en 2019, la plateforme **recense des schémas** partagés par différents producteurs de données et **permet d'accéder à la documentation** d'un schéma avec ses évolutions au cours du temps

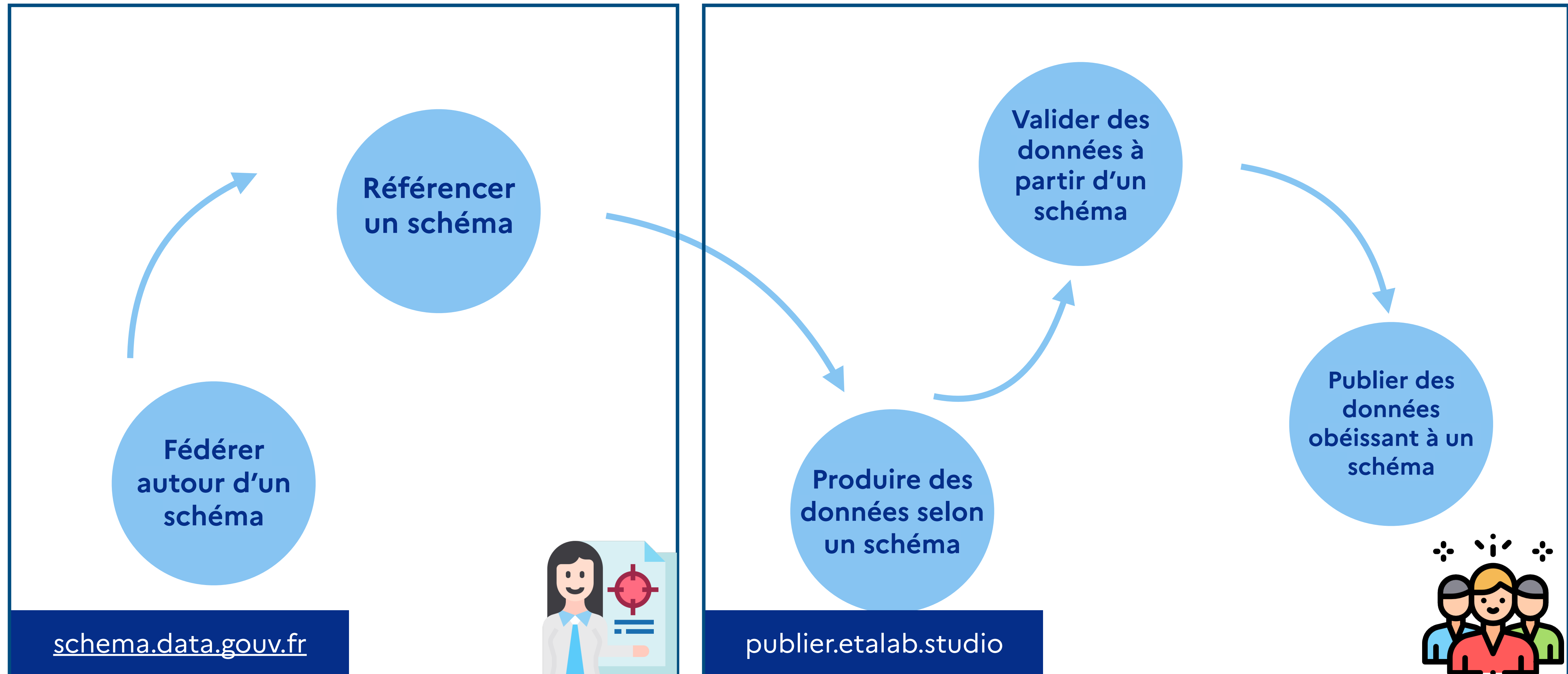
Objectifs :

- **Monter en qualité des données** en partageant entre différents producteurs un schéma unique de référence
- **Faire vivre une communauté** d'utilisateurs et proposer des guides méthodologiques pour faire adopter la démarche à différentes initiatives

Cycle de vie des données de qualité



Cycle de vie des données de qualité



Fonctionnalités publier.etalab.studio

- Valider la qualité d'un jeu de données
- Proposer un **parcours utilisateur sans rupture** entre les outils
- Faciliter la publication sur data.gouv.fr
- Fiabiliser la liaison entre un jeu de données et un « schéma »
- Proposer des **outils de saisie pour tout nouveau schéma référencé** sur schema.data.gouv.fr
- Faciliter la **consolidation** automatique de données issues d'un même schéma



publier.etalab.studio
Pour la publication de données de qualité

Geoffrey Aldebert ▾

Saisissez, validez et publiez vos données structurées

Vous utilisez un schéma de référence pour vos données ?
Validez et corrigez vos fichiers en quelques clics
et publiez-les directement sur data.gouv.fr !

Sélectionner un type de données ▾



Produire des données de qualité, en 4 étapes !

			
Je sélectionne le type de jeu de données adapté	Je saisis mes données ou je charge un fichier existant	Je vérifie si la structure de mes données est conforme	Je publie mon fichier validé sur data.gouv.fr

Merci

Contact : schema@data.gouv.fr

Guide : <https://guides.etalab.gouv.fr/producteurs-schemas/>

