## [Open Data] Template de readme + tags

**Faire une copie, ne pas compléter ce template**

| Suivi de modification / validation : rédaction d’un readme |
| --- |
| Fonction | Qui | Date |
| Porteur de la demande | Laurie Alla | initialisation du document le 23/02/24 |
| Directrice générale | Stéphanie Combes | validation le 09/03/24 |

Informations synthétiques

| Plan détaillé |
| --- |
| Objectif du document | Lister les différents éléments présents sur un format de readme standardisé pour la mise en Open Data de bases de données |
| Date de publication | À déterminer |
| Proposition d’une structure | 1. **Création de l’organisation** :
* Nom de l’organisation
* Sigle de l’organisation (facultatif)
* Numéro d’identification d’entreprise SIRET (facultatif)
* Description de l’organisation
* Site web (facultatif)
* Logo de l'organisation
1. **Publication d’un nouveau jeu de données** :
* Porteur/propriétaire
* Titre du jeu de données
* Sigle (facultatif)
* Description de la base de données :
	+ Objectifs et finalités de la base de données
	+ Contexte de création
	+ Public cible
	+ Résultats associés à la création de la base
* Présentation des données :
	+ Méthodologie de collecte et critères d’inclusion
	+ Choix des variables
	+ Description détaillée de chaque champ (nom, type, signification, valeurs possibles ou intervalle, unité de mesure)
	+ Structure de la base et rationnel des fichiers (si la base est volumineuse, elle scindée en fichiers)
	+ Lien de téléchargement d’un échantillon de données
* Problèmes connus (spécificités des données, erreurs, limites)
* Instruction/pré-requis d’utilisation de la base
* Support
* Contributions
* Crédits
* Formulaire traçabilité
* DOI
* Licence
* Maintenance
	+ Fréquence de mise à jour
	+ Date de la dernière mise à jour
* Mots clés
* Couverture temporelle
	+ Date de début
	+ Date de fin
* Couverture spatiale
	+ Zone géographique de la collecte des données
	+ Granularité de la couverture spatiale
* Autre
 |

##  Gabarit du readme

*La documentation pourra être renseignée dans ce template ou directement sur git dans le format markdown (format natif du read.me), selon les préférences et compétences git du porteur de projet.*

*Un exemple est rédigé dans les cases pour VisioMel à titre indicatif, pensez à bien créer une copie de ce document et ne pas écrire par-dessus*

*Pour information, des images peuvent être ajoutées dans les parties descriptives à condition qu’un* ***lien hypertexte vers l’image*** *soit aussi fourni. Les tableaux ne sont pas acceptés.*

# Création de l’organisation :

## *Avant de déposer un jeu de données, vous devrez créer votre organisation dans data.gouv.fr. Il faudra indiquer le nom de l’organisation à créer pour vérifier que celle-ci n’existe pas déjà dans data.gouv.fr.*

*Si l’organisation dépositaire a préalablement été créée sur data.gouv.fr, renseignez seulement son nom (en toutes lettres).*

## Nom de l’organisation :

|  |
| --- |

## Sigle de l’organisation (facultatif) :

|  |
| --- |

## Numéro d’identification d’entreprise SIRET (facultatif) :

|  |
| --- |

## Description de l’organisation :

|  |
| --- |

## Site web (facultatif) :

|  |
| --- |

## Logo de l'organisation :

|  |
| --- |

# Publication d’un nouveau jeu de données :

## *Dès lors que l’organisation est créée, vous pouvez publier du contenu, c'est-à -dire soit un nouveau jeu de données soit une nouvelle réutilisation. La suite du readme est consacrée à la publication d’un nouveau jeu de données.*

## Porteur/Propriétaire :

## *Précisez le nom de l’organisation propriétaire (écriture exacte et lien de l’organisation préalablement créée sur data.gouv.fr) ainsi que les noms des personnes référentes.*

## *Si la base de données est déposée au nom d’une ou plusieurs personne(s) morale(s) plutôt qu’au nom d’une organisation, précisez de la même manière le nom de la(les) personne(s) propriétaire(s).*

|  |
| --- |

## Titre du jeu de données :

*Donnez le titre de la base de données*

|  |
| --- |

## Sigle du jeu de données (facultatif) :

*Précisez le sigle de la base de données si existant*

|  |
| --- |

## Description de la base de données :

## *Cette section doit notamment comprendre les objectifs et finalités de la base de données, le contexte de création, le public cible, les résultats auxquels la base de données a permis d’aboutir et les mots clés.*

### *Objectifs et finalités initiales de la base de données :*

|  |
| --- |

### *Contexte de création :*

*Précisez si la base de données a été créée dans le cadre d’un Data Challenge, d’un partenariat ou autre.*

*Dans un contexte de Data Challenge ou de partenariat avec le Health Data Hub, vous êtes priés de le mentionner ici.*

|  |
| --- |

### *Public cible :*

|  |
| --- |

### *Résultats associés à la création de la base :*

*Précisez si des résultats (algorithmes, codes sources, documentation, publications scientifiques, etc.) sont associés à la base de données.*

*A titre d’exemple :*

* *Dans un contexte de Data Challenge, indiquez les algorithmes produits, les scores de performance associés, les éventuelles publications ou présentations ayant pu avoir lieu notamment lors d’une cérémonie de remises de prix et ajoutez les liens vers les ressources disponibles.*

|  |
| --- |

## Présentation des données :

*Élaborez sur la méthodologie de collecte et les critères d’inclusion des patients, le choix des variables, la source des données (dossiers médicaux, objets connectés, enquêtes, etc.) une description détaillée de chaque champ (nom, type, signification, valeurs possibles ou intervalle, unité de mesure), structure de la base et rationnel des fichiers (si la base de données est volumineuse, elle est scindée en fichiers), lien de téléchargement vers un échantillon de données.*

### *Méthodologie de collecte et critères d’inclusion :*

|  |
| --- |

### *Choix des variables :*

|  |
| --- |

### *Description détaillée de chaque champ (nom, type, signification, valeurs possibles ou intervalle, unité de mesure) :*

|  |
| --- |

### *Structure de la base et rationnel des fichiers (si la base est volumineuse et scindée en fichiers) :*

|  |
| --- |

### *Lien de téléchargement d’un échantillon de données :*

|  |
| --- |

## Problèmes connus :

*Indiquez s’il y a des spécificités, des points d’attention, des erreurs ou des limites à ce jeu de données*

|  |
| --- |

## Instructions/pré-requis d’utilisation de la base :

*Si des instructions ou des prérequis/précautions d’usage sont nécessaires, y compris en termes de pré-traitement des données, précisez les*

|  |
| --- |

##  Support :

*Indiquez si les utilisateurs peuvent obtenir un support et indiquez une adresse mail le cas échéant*

|  |
| --- |

## Contributions :

*Dites si vous acceptez les contributions et spécifiez vos exigences à cet égard.*

|  |
| --- |

## Crédits :

*Si d’autres ressources en lien avec cette base de données sont utiles à sa réutilisation (articles scientifiques ou toute autre ressource pouvant permettre une bonne réutilisation de la base de données), listez-les.*

|  |
| --- |

## Formulaire de traçabilité (facultatif) :

*Ajoutez un lien vers le formulaire à destination de l’utilisateur.*

*L’objectif du formulaire utilisateur est de pouvoir suivre qui (en termes de personnes et d’institutions) utilise les données et éventuellement à quelles fins. Ce formulaire est facultatif et n’est pas bloquant dans le sens où les demandes d’accès ne seront jamais refusées.*

|  |
| --- |

## DOI :

*Si existant ou attribué par l’équipe Open Science du Health Data Hub*

|  |
| --- |

## Licence :

*Précisez la ou les licences associées à la base de données.
Le Health Data Hub recommande l’utilisation conjointe des licences CC-CY 3.0 et Etalab 2.0.*

| * Creative Commons Attribution ([CC BY](https://opendefinition.org/licenses/cc-by/))
* Creative Commons Attribution Share-Alike ([CC BY-SA](http://www.opendefinition.org/licenses/cc-by-sa))
* Creative Commons CCZero ([CCZero](http://www.opendefinition.org/licenses/cc-zero))
* Licence Ouverte/Open Licence ([LO/OL](https://www.etalab.gouv.fr/licence-ouverte-open-licence))
* Licence Ouverte/Open Licence 2.0 ([LO/OL 2.0](https://www.etalab.gouv.fr/licence-ouverte-open-licence) ou Etalab 2.0)
* License not specified
* Open Data Commons Attribution Licence ([ODC-By](https://opendatacommons.org/licenses/by/summary/))
* Open Data Commons Open Database License ([ODbL](http://opendatacommons.org/licenses/odbl/summary/))
* Open Data Commons Public Domain Dedication and Licence ([PDDL](http://opendatacommons.org/licenses/pddl/summary/))
* Other (Attribution) : *Précisez*
* Other (Open) : *Précisez*
* Other (Public Domain) : *Précisez*
 |
| --- |

## Maintenance :

*Précisez si la base de données est maintenue et si oui, la fréquence des mises à jour*

### Fréquence de mise à jour :

| * Inconnu
* Ponctuelle
* Temps réel
* Toutes les heures
* Quatre fois par jour
* Trois fois par jour
* Deux fois par jour
* Quotidienne
* Quatre fois par semaine
* Trois fois par semaine
* Deux fois par semaine
* Hebdomadaire
* Toutes les deux semaines
* Trois fois par mois
* Deux fois par mois
* Mensuelle
* Bimestrielle
* Trimestrielle
* Trois fois par an
* Semestrielle
 |
| --- |

### Date de la dernière mise à jour :

| JJ/MM/AAAA |
| --- |

## Mots clés :

*Soyez exhaustif sur les mots clés à utiliser :*

* *Sigles et nom complet de l’organisation dépositaire, à titre d”exemple : “sfp” ; “Société Française de Pathologie” ; “sfd” ; “Société Française de Dermatologie” ; “gcc” ; “Groupe de Cancérologie Cutanée”*
* ***Health Data Hub avec les nomenclatures suivantes : “health-data-hub” ; “health data hub” ; “hdh” ;***
* *Si la base de données déposée à été utilisée à l’occasion d’un projet data challenge, ajoutez les tags suivants :* ***“data-challenge”; “data-challenges” ; “data challenge”***
* *Mots clés sur la thématique et les applications de la base de données, à titre d’exemple : “melanoma” ; “IA” ; artificial intelligence” ; “relapse” ; “prediction”*

|  |
| --- |

## Couverture temporelle :

*Précisez la date de début et la date de fin pour l’inclusion des données*

### Date de début :

| JJ/MM/AAAA |
| --- |

### Date de fin :

| JJ/MM/AAAA |
| --- |

## Couverture spatiale :

*Indiquez la zone géographique de la collecte des données et sa granularité*

### Zone géographique de la collecte des données :

|  |
| --- |

### Granularité spatiale :

| * Groupement de pays
* Pays
* Sous-ensemble de pays
* Région française
* Département français
* Arrondissement français
* Commune française
* Iris (quartiers INSEE)
* Canton français
* Collectivités d’outre-mer françaises
* Intercommunalité française (EPCI)
* Point d’Intérêt
* Autre
 |
| --- |

## Autre :

*Liens vers de la documentation ou autre*

|  |
| --- |

##

Tags/filtres inclus dans la fonction de recherche

| **Catégories** | **Tags possibles (non-exhaustif)** |
| --- | --- |
| **Mots clés** | *Texte libre* |
| **Formats** | * csv
* txt
* zip
* json
* pdf
* xls
* xlsx
* tiff
* jpeg
* jpg
* png
* Autre : *préciser*
 |
| **Licences** | * Toutes les licences
* Creative Commons Attribution
* Creative Commons Attribution Share-Alike
* Creative Commons CCZero
* Licence Ouverte/Open Licence
* Licence Ouverte/Open Licence 2.0 (Etalab 2.0)
* License not specified
* Open Data Commons Attribution Licence
* Open Data Commons Open Database License (ODbL)
* Open Data Commons Public Domain Dedication and Licence (PDDL)
* Other (Attribution)
* Other (Open)
* Other (Public Domain)
 |
| **Schéma** | * Tous les schémas
 |
| **Couverture spatiale**  |  *Texte libre*  |
| **Granularité spatiale** | * Groupement de pays
* Pays
* Sous-ensemble de pays
* Région française
* Département français
* Arrondissement français
* Commune française
* Iris (quartiers INSEE)
* Canton français
* Collectivités d’outre-mer françaises
* Intercommunalité française (EPCI)
* Point d’Intérêt
* Autre
 |

##