



HEALTH DATA HUB

Données de santé, open source et science ouverte

Meetup #6 du Health Data Hub
26.11.2020



BUT

Créé par la loi de modernisation du système de santé de 2016, le Système National des Données de Santé (SNDS) collecte et met à disposition, à des fins définies, des données de santé individuelles pseudonymisées provenant des principales sources médico-administratives.



LE SNDS HISTORIQUE



Données de l'assurance maladie



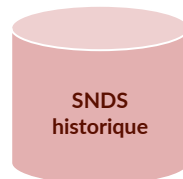
Données des hôpitaux



Causes des décès



Alimentation du SNDS historique



PÉRIMÈTRE ÉLARGI

- Jusqu'en 2019, le SNDS historique était limité aux données des demandes de remboursement de l'assurance maladie, aux données des sorties d'hôpital et aux causes de décès (DCIR, PMSI et CépiDC)
- Depuis 2019, par le biais de l'article 41 de la [loi sur l'organisation et la transformation du système de santé](#), le champ d'application a été élargi



CHIFFRES

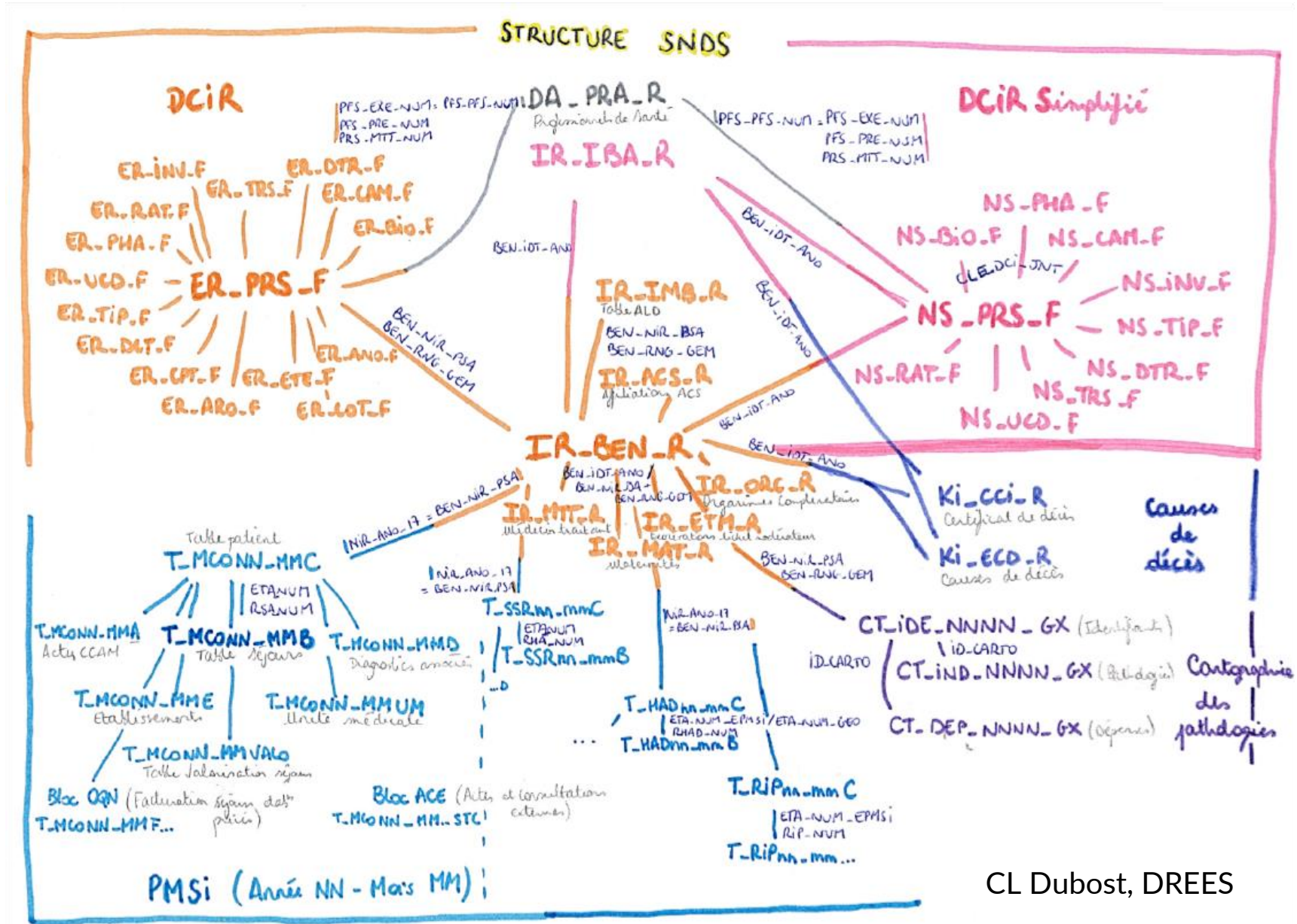
Une des plus grandes base de données médico-administratives au monde

Données sur les remboursements de 65 millions de personnes
1,2 milliard de feuilles de soins par an
11 millions de séjours hospitaliers par an

Le Système national de données de santé historique (SNDS historique)

- +180 tables
- +4500 variables
- Nombreuses règles métiers
- Évolution temporelle

=> L'accès aux données ne suffit pas pour les utiliser.



CL Dubost, DREES

VERROU



ACCÈS AUX DONNÉES

Complexité réglementaire



UTILISATION DES DONNÉES

Complexité du schéma des données du SNDS
historique

Complexité dans la manipulation des données, biais

Complexité dans l'interprétation des données

SOLUTIONS PROPOSÉES



ACCÈS AUX DONNÉES

Document pédagogique facilitant la constitution des dossiers de demandes d'autorisation CNIL



UTILISATION DES DONNÉES

Dictionnaire des tables et variables et de leurs relations

Documentation collaborative ouverte

Données synthétiques pour visualiser les données et arbitrer la pertinence d'une demande d'accès

Forum d'entraide pour échanger avec des experts

L'open source et science ouverte au sein de l'écosystème de la donnée de santé



Le Health Data Hub a pour mission de favoriser le partage, sous licence ouverte, des **connaissances** et **ressources** sur le **SNDS historique** et **autres bases de données**. **Différents contributeurs** et **utilisateurs** sont à l'origine de ce partage et participent à la dynamique.

WP 4: Standardisation



WP 1: Gouvernance et partenariat



WP 3: Animation, formation, support



WP 2: Outillage et maintenance



Les outils ouverts mis en place



Vision du Health Data Hub à propos de l'open source



QUELQUES CHIFFRES (SEPTEMBRE 2020)

1033 membres dans la communauté Meetup

22 évènements organisés pour la communauté sur Meetup depuis mars 2019

4 ressources de connaissances (documentation collaborative, dictionnaire interactif, librairie de programmes et jeux de données de synthèse)

3 notebooks (en R et python) pour se former au SNDS historique (sur deux jeux de données de synthèse différents)

2 outils pour collaborer (Gitlab, Forum Discourse)



OBJECTIFS DANS 1 AN

1500+ membres qui suivent l'actualité du Health Data Hub sur Meetup

Une série d'évènements en visioconférence sur une thématique choisie organisés sur Meetup

Un toolkit de ressources pour une prise en main facilitée du SNDS historique

Une librairie de programmes sur les concepts de base sur le SNDS historique

Déclinaison des ressources de connaissance aux premières bases disponibles via le catalogue du HDH

CONTRIBUTEURS



<https://documentation-snds.health-data-hub.fr/>



Zoom sur la documentation collaborative

Documentation du SNDS

Forum entraide Dico interactif Groupe Meetup GitLab

Introduction ▶

Fiches ▶

Glossaire ▶

Se former au SNDS ▶

Pour aller plus loin ▶

Évènements passés et à venir ▶

Tables ▶

Contribuer ▶

À propos ▶

Section introduction



Système national des données de santé

Cette section héberge un guide introductif au SNDS, décliné en quatre sections :

- [Qu'est-ce que le SNDS ?](#)
- [Les bases du SNDS, de l'alimentation à la restitution des données](#)
- [Comment accéder au SNDS ?](#)
- [La recherche avec le SNDS](#)

Les éléments spécifiques sont réservés à la section [Fiches thématiques](#).

Attention

La loi du 24 juillet 2019, relative à l'Organisation et la Transformation du Système de Santé (OTSS), modifie certains éléments présentés ici, en particulier en élargissant le périmètre du SNDS et en précisant les missions du Health Data Hub. La loi n'étant pas encore entièrement mise en application, nous nous concentrerons ici sur le SNDS "historique", déjà complexe, et

SE FORMER AU SNDS

Kit de démarrage avec les données de santé

Guides et formations de la CNAM

“SNDS, ce qu’il faut savoir” de Santé publique France

Données de synthèse

3 notebooks (en R et python) pour se former au SNDS historique

Une dizaine de programmes



opensource@health-data-hub.fr



HEALTH DATA HUB

Notebooks - Prise en main du SNDS à partir d'exercices simples et guidés sur les données synthétiques

Meetup #6 du Health Data Hub
26.11.2020



Qu'est-ce qu'un Notebook ?

Un notebook est un **cahier électronique** qui, dans un même document, peut rassembler **du texte et du code informatique exécutable**. Les notebooks développés par le HDH vous accompagnent dans la **découverte de quelques tables majeures du SNDS historique à partir d'exercices simples et guidés**. Ils sont consultables/téléchargeables sur le [Gitlab du HDH](#).



Public concerné

Les notebooks se destinent aux **primo-utilisateurs du SNDS historiques**. Ils permettent de gagner une **première sensation des données** disponibles dans le SNDS historique et de leur structure (en attendant d'avoir accès aux données réelles) grâce à la manipulation de **données fictives**.



Thématiques abordées

Actuellement, trois notebooks sont disponibles :

- **cohort** : constitution d'une cohorte à partir de critères particuliers, consommation de soins de ville, hospitalisations
- **pmsi** : indicateurs hospitaliers dans le PMSI MCO
- **diagnostics** : diagnostics hospitaliers dans le PMSI MCO



Données utilisées

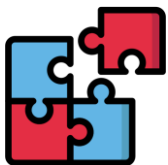
Les notebooks reposent sur les **données de synthèse du SNDS historique**. Chaque notebook est disponible en 2 versions : celui pour les **données du Lab Santé de la DREES** (disponible en open data) et celui pour les **données de REPERES** (accès via plateforme dédiée). Pour plus d'informations sur les données de synthèse, cf. la [documentation en ligne](#).



Langages informatiques

Chaque notebook est disponible en 2 langages : un en **Python** et un en **R**. Il s'agit de deux langages informatiques disponibles sur la plateforme technologique du HDH.

Points d'attention



Incomplets

Certains **filtres de “nettoyage”** pouvant être réalisés sur le SNDS historiques sont **volontairement manquants** pour rendre la prise en main plus aisée.



Simplifiés

Les approches mises en oeuvre dans les notebooks sont **simplifiées par rapport à la réalité** afin d'aller à l'essentiel. Il s'agit d'outils pédagogiques et en aucun cas de guides méthodologiques pour des études sur le SNDS historique.



Adaptés

Le code a été **adapté à la structure des données fictives** qui présente des limites et ne reflète pas nécessairement celle du vrai SNDS historique (format des variables, temporalité des données, liens relationnels, etc.).



Non optimisés

Le code **n'a pas été optimisé pour gagner en temps de requête** ; cela peut s'avérer nécessaire lorsqu'on travaille sur le SNDS historique dont la volumétrie est beaucoup plus importante que les données de synthèse.



Le code des exercices ne peut être utilisé tel quel et nécessitera d'être retravaillé pour répondre à la réalité du SNDS historique



Les données fictives utilisées ne sont pas nécessairement plausibles et ne recouvrent pas une réalité médicale

Démonstration - Notebooks Lab Santé

1. Télécharger les notebooks (fichiers au format .ipynb) sur le Gitlab du HDH

Vous rendre ici : <https://gitlab.com/healthdatahub/formation/exercices-snds/-/tree/master/notebooks>

Consulter/Télécharger le notebook souhaité : Notebook_ *thématique_langage* _Lab

2. Ouvrir le fichier téléchargé dans un environnement dédié

Sur internet : [Google collaboratory](#), [JupyterLab en ligne](#), etc.



Sur votre poste : Anaconda (gratuit), etc.



3. Exercez-vous sur le notebook

Exercices de découverte du SNDS (Version R)

Bienvenue sur ce notebook d'introduction à la manipulation du SNDS ! L'objectif est de vous accompagner, grâce à une manipulation guidée, dans la découverte de quelques tables majeures du SNDS. Ces quelques exercices sont conçus pour vous donner une première prise en main des données disponibles et de leur structure, en attendant que vous ayez accès aux données réelles dont vous avez besoin pour vos projets.

Nous allons ici constituer une cohorte de patients répondant à des critères particuliers, et nous intéresser à leur consommation de soins de ville (grâce au DCIR) et à l'hôpital (grâce au PMSI) au fil des différents exercices.

A noter que les données utilisées pour le développement de ce notebook sont des données synthétiques avec les mêmes formats et types des valeurs des données réelles du SNDS. Il s'agit de données fictives générées aléatoirement, qui ne sont pas nécessairement plausibles et ne recouvrent pas une réalité médicale. Ces données ont été générées par le Lab Santé de la DREES.

Pour plus d'information sur ces données, vous pouvez consulter la [page dédiée sur le site de documentation du SNDS](#).

Préparation et configuration

Commençons par préparer le terrain. Nous utiliserons dans les exercices un certain nombre de packages pour nous faciliter la tâche. Vous devrez ainsi avoir installé les modules suivants pour exécuter le code :

reshape2, ggplot2, dplyr et data.table (voir le code pour les détails).

```
[ ]: # Installation des packages (cela prends du temps)
#install.packages("ggplot2")
#install.packages("dplyr")
#install.packages("reshape2")
install.packages("data.table")

library(reshape2) # Manipulation de données
library(data.table) # Manipulation de données
library(dplyr) # Manipulation de données
library(ggplot2) # Visualisation de données

Installing package into '/usr/local/lib/R/site-library'
(as 'lib' is unspecified)
```

Pour information :

- Les bases de données du Lab Santé utilisées sont disponibles sous .csv [ici](#)
- Une [fiche "Notebooks"](#) est disponible sur la documentation SNDS collaborative

Remerciements

Pour leur relecture et idées d'amélioration des notebooks :



Céline Leroy (ARS Normandie)
Jérôme Brocca (ARS Centre-Val de Loire)



Philéas Condemine (Drees)



Laure Fernandez de Martini (CNAM)
Julien Brand (CNAM)



Health Data Hub

FAQ SNDS : <https://entraide.health-data-hub.fr/>