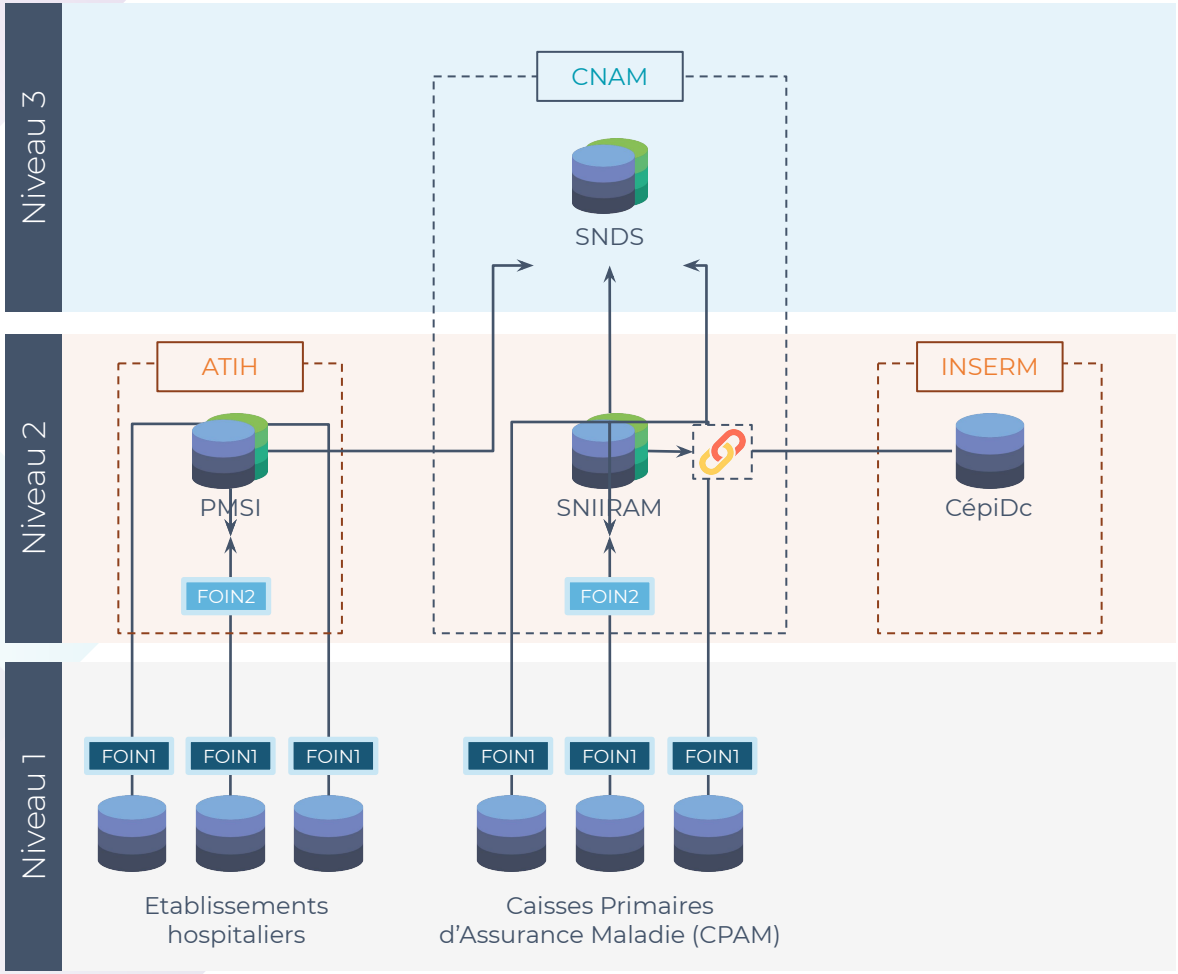
HEALTH DATA HUB

SÉCURITÉ SOCIALE
l'Assurance Maladie
Caisse Nationale

Schémas d'appariement et de circulation des données

1. Rappel sur le SNDS

Rappel sur le SNDS - périmètre historique (1/2)



Base de données sources
Base de données consolidée
Appariement entre deux bases

Le SNDS est alimenté par trois bases de données :

- SNIIRAM

  Les Caisses Primaires d'Assurance Maladie remontent l'ensemble des informations issues des remboursements à la CNAM. Pour garantir la sécurité de ces données, la solution CAMEL est utilisée permettant le contrôle, le chiffrement, et le transfert sécurisé des flux vers le SNIIRAM. Le flux de données provenant des caisses passe par CAMEL qui applique la fonction de hachage FOIN1 sur le NIR.. Les données sont envoyées à la CNAM qui applique ensuite l'algorithme de hachage FOIN2.

- PMSI

  Chaque établissement de santé enregistre chacun des séjours hospitaliers, appelé résumé de sortie standardisé (RSS) qui sont classés dans des Groupe Homogène de Malade (GHM) selon des critères médicaux et économiques. Les établissements utilisent alors le logiciel MAGIC qui applique la fonction de hachage FOIN1 sur le NIR et dépose les données pseudonymisées sur la plateforme e-PMSI de l'ATIH.

  L'ATIH applique alors l'algorithme de hachage FOIN2 et envoie les données à la CNAM pour intégration dans le SNDS.

- CépiDc

  Le CépiDc de l'Inserm gère la Base de Causes Médicales de Décès (BCMD). Il ne traite que la partie médicale du certificat de décès, par conséquent, la base ne contient aucun nom. Les données sont transmises à la CNAM pour intégration dans le SNDS et sont appariées de façon probabiliste au SNIIRAM pour récupérer les FOIN2 correspondants afin de lier les causes de décès aux autres données présentes dans le SNDS.

Rappel sur le SNDS - périmètre historique (2/2)

Il couvre l'ensemble des personnes ayant eu **recours au système de soins français ou étant décédées sur le territoire.** Il permet de disposer d'informations sur le recours aux soins, le montant des dépenses associées, les informations concernant les **séjours hospitaliers depuis 2006**, ainsi que les **causes de décès des individus sur la période 2013-2015**.

Les **principales informations du** SNDS sont :

- Informations sur le bénéficiaire (sexe, mois et année de naissance, rang de naissance, lieu de résidence, régime, couverture maladie universelle complémentaire, aide à la complémentaire santé) ;
- Informations sur les professionnels de santé (spécialité, mode d'exercice, sexe, âge, département d'implantation)
- Pathologies, notamment les affections de longue durée et les diagnostics des séjours hospitaliers ;
- Dépenses et remboursements (prestations en soins de ville, en établissements de santé, et montants associés)
- Consommations de soins de ville (consultations, actes techniques…)
- Prescriptions (médicaments)
- Dispositifs médicaux (aides techniques)
- Autres prestations (cures, transports…)
- Soins hospitaliers (hors séances)
- Séjours hospitaliers
- Indemnités journalières (maladie, ATMP, maternité) et invalidité
- Causes médicales de décès

Pour approfondir

- Dictionnaire des variables : lien
- Documentation du SNDS : lien
- Forum d'entraide : lien
- Contribuer (Gitlab) : lien
- Programme de formation – à venir
- Note sur l'impact de la loi OTSS – à venir

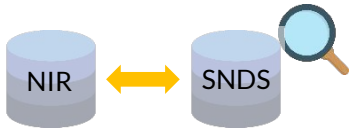
2. Circulation et appariement des données

Présentation des différents types d'appariement

1. Appariements déterministes directs (NIR)

Couplage de données réalisé à partir du NIR

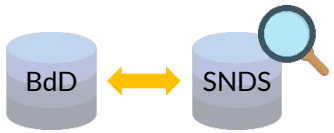
Appariement déterministe direct (NIR) entre une base de données et le SNDS



2. Appariements indirects

Couplage statistique de données réalisé à partir d'un ensemble de variables discriminantes et communes aux deux jeux de données

Appariement indirect entre une base de données et le SNDS



Principaux types d'appariement détaillés ci-après.
En cas de questions, n'hésitez pas à consulter le forum d'entraide : https://entraide.health-data-hub.fr/ ou à contacter le HDH : contact@health-data-hub.fr

Appariement direct (NIR) entre une base de données et le SNDS (projet hébergé à la CNAM ou au HDH)

#1 Extraction des données de la cohorte

1. Le producteur réalise une extraction de sa base source, avec :
   - L'ID interne : ID_source
   - L'identifiant SNDS en clair : NIR de l'ouvreur de droit, sexe et date de naissance du bénéficiaire et/ou NIR du bénéficiaire
   - Les données métier dé-identifiées
2. Un nouvel ID est généré aléatoirement : ID_temp
3. Une table de pseudonymisation [ID_temp ; ID_SNDS_clair] est envoyée dans le circuit de pseudonymisation SNDS via la procédure SAFE https (maîtrisée par la CNAM)
4. L'extraction de la base source est envoyée au HDH ou à la CNAM avec :
   - ID_temp
   - Les données métier dé-identifiées

Producteur

1. Extraction de la base source
2. Nouvel ID
3. Transmission du fichier procédure SAFE
4. A livrer au HDH/CNAM

Pseudonymisation des NIR

#2 Pseudonymisation de l'identifiant SNDS

1. La procédure SAFE https (maîtrisée par la CNAM) reçoit en entrée la table de pseudonymisation avec :
   - ID_temp
   - ID_SNDS_clair
2. L'identifiant SNDS en clair est pseudonymisé de manière irréversible par les algorithmes de hachage FOIN : ID_SNDS_clair est transformé en ID_SNDS_pseudo
3. En sortie de la procédure SAFE https, la table des identifiants pseudonymisés est reçue par la CNAM, avec :
   - ID_temp
   - ID_SNDS_pseudo

CNAM

1. Transmission par le producteur
2. Hachage de ID_SNDS_clair
3. Reçu à la CNAM

Légende

- Identifiants
  - ID_source
  - ID_temp
  - ID_projet
  - Aléatoires
- NIR
  - ID_SNDS_clair
  - ID_SNDS_pseudo
- Données
  - Données métier
  - Données SNDS

Chaînage des données du SNDS avec l'identifiant SNDS

#3 Extraction des données SNDS

1. La CNAM reçoit via la procédure SAFE https la table des identifiants pseudonymisés, avec :
   - ID_temp
   - ID_SNDS_pseudo
2. La CNAM cible les données du SNDS requises via l'ID_SNDS_pseudo
3. Préparation de la mise à disposition (hébergement à la CNAM ou au HDH) :
   - ID_temp
   - Données du SNDS requises (sans ID_SNDS_pseudo)

CNAM

1. Reçu SAFE
2. Extraction SNDS
3. Préparation mise à disposition

Livraison des données SNDS

#4 Mise à disposition des données

1. Le HDH ou la CNAM apparie les extractions transmises par la source métier et l'extraction SNDS à partir de l'identifiant ID_temp.
2. Un nouvel ID est généré aléatoirement : ID_projet
3. Un jeu de données final est livré à l'équipe projet, avec :
   - ID_projet
   - Les données métier (dé-identifiées)
   - Les données SNDS (dé-identifiées)

CNAM/HDH

1. Appariement des extractions reçues depuis les sources
2. Nouvel ID
3. Livraison au porteur

Livraison des données métier

HEALTH DATA HUB

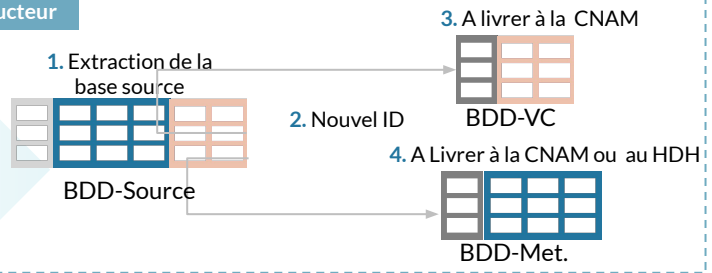
L'Assurance Maladie – Caisse Nationale

Appariement indirect entre une base de données et le SNDS (projet hébergé à la CNAM ou au HDH)

#1 Extraction des données de la cohorte

1. Le producteur réalise une extraction de sa base source, avec :
   - L'ID interne : ID_source
   - Les variables communes
   - Les données métier **dé-identifiées** (Met.)
2. Un nouvel ID est généré aléatoirement par le producteur de données : ID_temp
3. Une table de variables communes est envoyée à la CNAM avec :
   - ID_temp
   - Les variables communes (VC)
4. L'extraction de la base source est envoyée au HDH ou à la CNAM avec :
   - ID_temp
   - Les données métier **dé-identifiées** (Met.)

Producteur



Livraison de la table de variables communes (BDD-VC) via PETRA

#2 Extraction des données SNDS

1. La CNAM reçoit du producteur la table BDD-VC de variables communes avec :
   - ID_temp
   - Les variables communes
2. À partir des variables communes, la CNAM effectue une recherche des patients dans le SNDS en ciblant les données pertinentes.
3. L'extraction du SNDS est envoyée au HDH ou la CNAM, avec :
   - ID_temp (pour lesquels l'appariement est validé)
   - Données du SNDS requises (dé-identifiées), comprenant les variables communes

CNAM

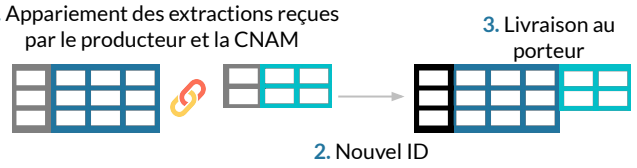
Livraison des données métier

Livraison des données SNDS

#3 Mise à disposition des données

1. Le HDH ou la CNAM apparie les extractions transmises par le producteur et la CNAM à partir de l'identifiant ID_temp.
2. Un nouvel ID est généré aléatoirement : ID_projet
3. Un jeu de données final est livré à l'équipe projet, avec :
   - ID_projet
   - Les données de métier dé-identifiées
   - Les données SNDS (dé-identifiées)

CNAM/HDH



Légende

Identifiants
- ID_source
- ID_temp
- ID_projet

} Aléatoires

NIR
- ID_SNDS (NIR foinisé)

Données
- Données métier
- Données SNDS
- Variables communes

Suivez-nous sur les réseaux sociaux !