



HEALTH
DATA HUB



Les chaînages de données

conception et expertise

Sanoia
Real World Digital CRO

Un document qui s'inscrit dans un objectif de formation citoyenne



Ce document vise à informer le public sur les usages pour la recherche avec les données du **Système National des Données de Santé** (SNDS). Mieux connaître le cadre et les enjeux liés aux données de santé, et notamment du SNDS, en partant des questionnements des citoyens.

Une initiative de :



Modules

- Les données dans le SNDS et de la démarche de recherche
- L'accès aux données du SNDS
- Ressources pour les associations
- Les chaînages de données
- Un projet avec chaînage - HYDRO
- Intelligence Artificielle en santé
- La démarche Open Data
- Projet de partage - Vivre-Covid19



Illimité



Tout public



20 min



Support en accès libre



Webinaires *Les Mardis de la Donnée de Santé*

Ref. SNDS CITY - 001

Les contenus de ce module ont été développés avec le soutien de :



Pourquoi informer et former sur les données de santé ?



Pour permettre un système de santé efficace, personnalisé et humaniste, les données de santé ont un rôle à jouer. C'est par le traitement et le croisement d'un grand volume de données de qualité, que les recherches les plus impactantes pourront être menées : améliorer le dépistage et le diagnostic d'une maladie, analyser les effets secondaires des traitements, faire évoluer les essais cliniques.



Les données représentent un patrimoine de grande utilité pour la société. Elles sont, entre autres, issues de la Caisse Nationale d'Assurance Maladie, des établissements de santé, de l'Institut National du Cancer, de Santé Publique France, de l'Agence de Biomédecine, de différentes sociétés de médecine.



Le Health Data Hub et ses partenaires, proposent une série de modules de formation aux données de santé, afin que la société civile dispose de la connaissance nécessaire et puisse ainsi être informée de façon compréhensible sur les projets liés aux données de santé, sur les droits dont chacun et chacune dispose vis à vis de ces données, voire que ses membres s'engagent dans un projet avec les données...

« Donner son sang, c'est aider son prochain ;
il en est de même du partage de ses données
de santé. Les partager à son équipe
soignante, c'est être mieux soigné ; les
partager pour tous, c'est permettre que
l'ensemble de nos concitoyens le soient aussi.
»



Gérard Raymond
Président de France Assos Santé
Vice-président du Health Data Hub

Nous vous proposons de...

Mieux connaître le cadre et les enjeux liés à la réutilisation des données de santé à finalité de recherche -et notamment du SNDS* - et plus largement des technologies associées.

Nous commençons par découvrir la base principale du Système National des Données de Santé (SNDS*) et ce qu'elle contient (module 1) ;

Avant tout, il est utile de décrire quels acteurs peuvent accéder au SNDS (module 2) ;

Vous suivrez alors ensuite une association de patients qui monte un projet d'utilisation des données de santé pour voir les bonnes pratiques (module 3) ;

Mais il n'existe pas que la base principale du SNDS ! A travers l'exemple d'une base construite par un tiers, on explore les possibilités de chaînage des bases de données existantes avec le SNDS (module 4) ;

En pratique, ça marche ! Vous verrez par exemple comment s'est mis en place un projet en Intelligence Artificielle, le projet HYDRO. Ce type de projet "hybride" combine les données du SNDS avec une base externe pour innover dans le domaine des dispositifs médicaux (module 5) ;

A cette fin, le Health Data Hub propose une plateforme tierce de croisement, facilitant en particulier la réalisation de projets d'Intelligence Artificielle (module 6) ;

En parallèle, nous explorerons les données "open-data" qui, bien plus qu'une alternative, peuvent ouvrir des possibilités accrues de partage (module 7) ;

Mais pour tout ça, il faut ... des données ! Ainsi vous pourrez disposer d'un retour d'expérience d'un projet de création d'une base de données en direct des citoyens : le projet Vivre-Covid19 (module 8).

Qu'allons-nous apprendre aujourd'hui ?

Dans ce module :

Nous suivons une équipe qui utilise des données existantes pour produire de nouveaux résultats

Lorsque nous les suivons, nous répondrons avec eux aux problématiques qu'ils rencontrent :

1. L'équipe souhaite **utiliser une base de données de santé existante** : dans quelles mesures (réglementaires et opérationnelles) pourrait-elle le faire ?
2. Dans un premier temps, elle pourrait utiliser des **données anonymes en accès libre**.
3. Si cette solution ne convient pas, alors l'équipe peut demander à **utiliser des données plus complètes** voire **les compléter avec d'autres bases de données de recherche en santé** en respectant un cadre réglementaire strict.
4. **Quelles données sont disponibles dans le SNDS** afin de compléter leur étude ?
5. Et comment est-ce qu'elle peut les utiliser pour **compléter sa base de données en pratique** ?
6. Quelles sont les démarches juridiques à suivre ? **Dans quel cadre leur étude se place-t-elle** ?



Ainsi en sortant de ce module, vous aurez les bases pour répondre à toutes les questions ci-dessus, même dans d'autres contextes !

Plan

- 1. Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane**
- 2. L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?**
- 3. Comment fait cette équipe en pratique pour compléter la base du projet Diane ?**
- 4. Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.**
- 5. Et si la base Diane était encore en cours de constitution ? Ce serait différent ?**
- 6. Quelles seront les nouvelles possibilités de demain ?**

Plan

1. **Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane**
2. L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?
3. Comment fait cette équipe en pratique pour compléter la base du projet Diane ?
4. Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.
5. Et si la base Diane était encore en cours de constitution ? Ce serait différent ?

Une équipe du CHU de Nice veut monter un projet de recherche...



Voici Emma, médecin endocrinologue au sein du centre hospitalier universitaire (CHU) de Nice.

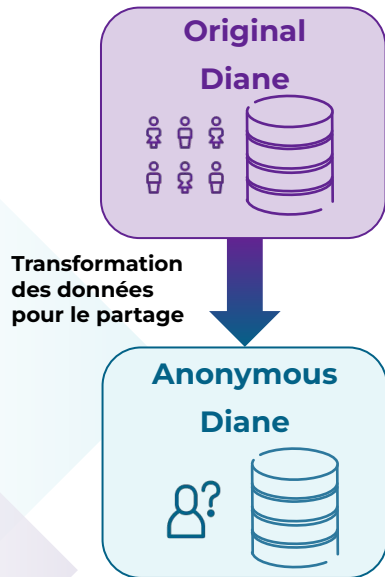
Depuis quelque temps, **elle s'aperçoit que beaucoup de ses patients diabétiques ont eu recours à un acte médical particulier dans les 2 ans avant le début de leurs premiers symptômes**. Cela l'interpelle et elle s'interroge sur la possible généralité de cette pratique : **est-ce qu'il y aurait des événements particuliers dans leurs historiques médicaux qui préfigurent le début de leur maladie ?**

Elle peut vérifier cela à son échelle, en regardant les dossiers médicaux de ses patients mais elle aimerait vérifier ses résultats à plus grande échelle.

Pour ce faire, **elle veut monter un projet de recherche avec une équipe de son CHU et des collègues intéressés dans d'autres hôpitaux.**



... Ils connaissent le projet Diane qui s'est terminé l'an dernier...



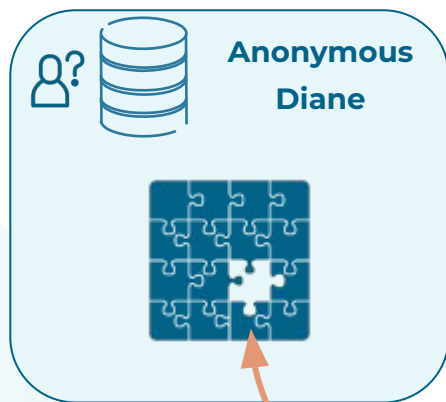
Lors des discussions préliminaires au sein de l'équipe, ils ont remarqué qu'il existait déjà une cohorte (= groupe de patients qui participent à un projet de recherche) qui traitait un sujet similaire et qui s'est terminée l'an dernier : **le projet Diane a suivi la santé d'une cohorte de français au cours de la première année après le diagnostic de leur diabète entre 2010 et 2015.**

Les responsables de cette étude l'ont conçue de telle sorte qu'elle puisse être partagée avec d'autres équipes en cas de besoin.

En effet, sur le site web dédié de l'étude, il est possible de télécharger un **jeu de données anonyme** de la base mis en accès libre : c'est une version qui **ne permet pas de remonter aux individus qui la composent*** (agrégation, ajout de bruit, etc.)

➤ Pour commencer, l'équipe du CHU de Nice utilise la version anonyme de la base du projet Diane : peut-être contient-elle déjà toutes les données nécessaires pour répondre à leur question de recherche ?

... mais la base Diane ne contient pas assez d'informations : l'équipe veut donc la compléter.



Malheureusement, **le projet Diane ne collecte pas toutes les informations nécessaires** pour répondre à la question de recherche posée par l'équipe du CHU de Nice : **il n'interroge pas l'historique systématique des actes médicaux -avant le début de la maladie-** des patients inclus.

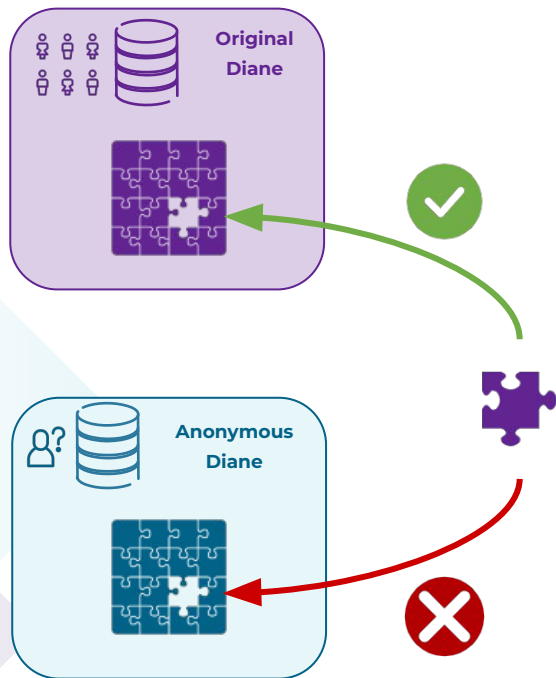
Néanmoins, les informations collectées par le projet Diane sont très intéressantes pour l'équipe du CHU de Nice.



Ainsi, plutôt que de repartir de zéro et constituer leur propre base de données, l'équipe voudrait **rajouter les informations manquantes nécessaires dans la base du projet Diane** : ils veulent créer la base **Diane 2.0**.

Mais alors, comment faire ?

... mais la base Diane ne contient pas assez d'informations : l'équipe veut donc la compléter.



Il faut rajouter des informations pour chaque patient de la base Diane : c'est **impossible dans une base anonyme**. L'équipe de Nice doit donc demander une extraction de la base Diane originale (c'est-à-dire celle ayant servi à la production de la base anonymisée) contenant **suffisamment d'informations au niveau individuel pour pouvoir retrouver chaque patient** dans la base de données contenant les informations supplémentaires nécessaires.

C'est simple en principe mais il y a 3 points importants à considérer :

- 🚩 **Où trouver les informations supplémentaires** : ici, ce sont des données médico-administratives donc elles seront répertoriées dans le SNDS.
- 🚩 **Comment compléter la base Diane** : on veut utiliser des données contenues dans le SNDS.
- 🚩 On n'utilise plus des données anonymes de santé mais des données personnelles de santé : **le cadre réglementaire est beaucoup plus strict.**

Plan

1. Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane
2. **L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?**
3. Comment fait cette équipe en pratique pour compléter la base du projet Diane ?
4. Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.
5. Et si la base Diane était encore en cours de constitution ? Ce serait différent ?

Quelles données sont présentes dans le Système National de Données de Santé (SNDS) ?

UN PRINCIPE FONDAMENTAL

Les données associées à chaque **remboursement** de l'Assurance Maladie (consultation, délivrance de médicament, biologie, imagerie, hospitalisation, etc.) sont **présentes dans le SNDS.**

MAIS par exemple, cela ne contient pas :

- le "résultat" de biologie suite à une prise de sang
- des données sur le mode de vie des personnes comme le tabagisme, la consommation d'alcool...
- les diagnostics effectués par un médecin généraliste ou spécialiste, en ville
- ...

Comment est construit le SNDS ? (base principale)

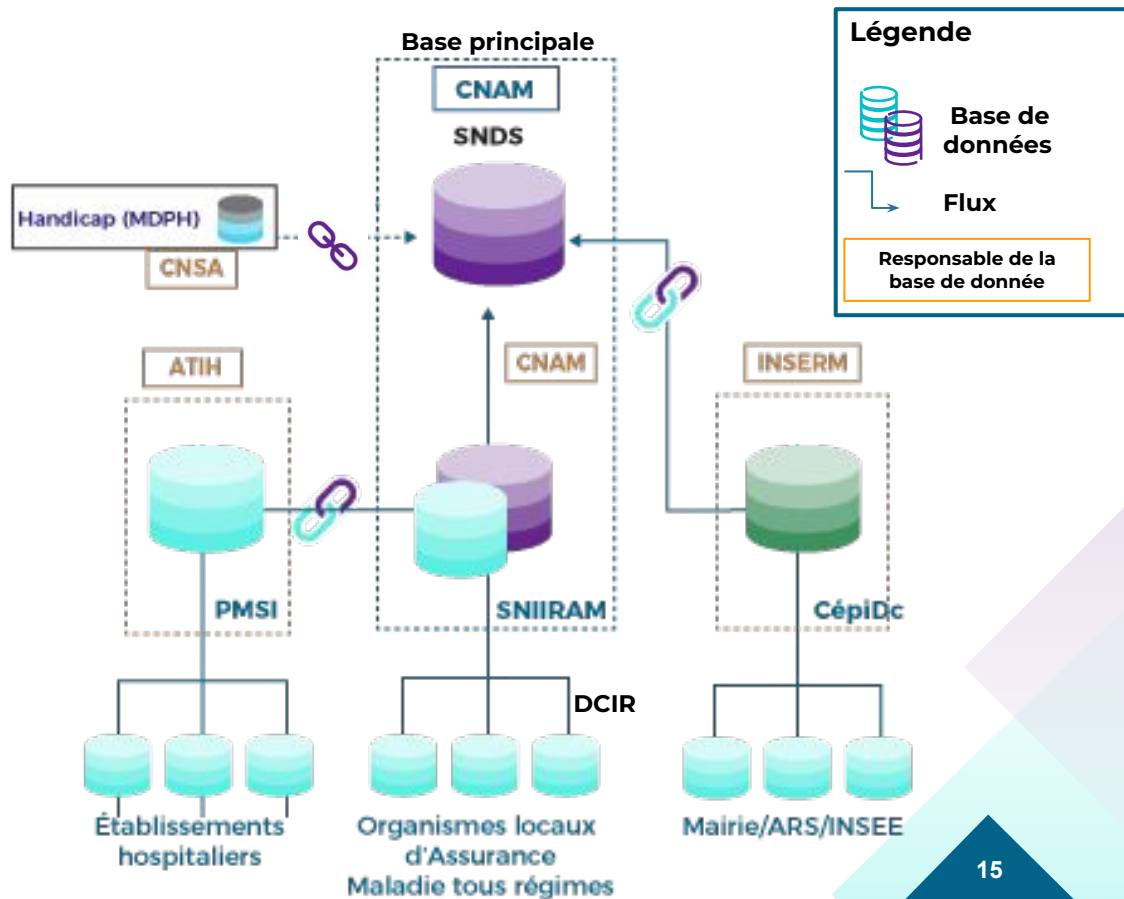
La **base principale du SNDS** regroupe des bases publiques de données de santé dont le PMSI, le SNIIRAM, CépiDc ou encore la base SI-DEP (non représentée)...

Elle est alimentée et **mise à jour régulièrement** par des flux de données issus de ces différentes bases.

L'accès au SNDS est strictement encadré, comme nous le verrons par la suite.



La base principale du SNDS est un ensemble **de bases pré-existantes** alimentées par différents circuits de données.



Quelles données sont disponibles dans le SNDS ? (1/3)

Le DCIR (Datamart de Consommation Inter Régime) contient l'ensemble des informations issues des remboursements de l'Assurance Maladie et est géré par la Caisse Nationale d'Assurance Maladie (CNAM).

≈1,2 milliard de feuilles de soins / an

Consommation des soins de ville

Données disponibles :

- Prestations remboursées : actes médicaux, biologie, dispositifs médicaux, médicaments, transports médicaux, soins dentaires, arrêts maladie, etc. ;
- Montants présentés au remboursement, montants remboursés par l'Assurance Maladie ;
- Dates de prescription, exécution et remboursement de soins.

Données démographiques des bénéficiaires

Données disponibles :

- Age, sexe, commune de résidence, date de décès ;
- Bénéficiaire de la CMU-C (Couverture Maladie Universelle Complémentaire) ou, depuis 2019, de la C2S (Complémentaire Santé Solidaire) ;
- Diagnostic de l'affection de longue durée.

Infos sur les prescripteurs / établissements et exécutants

Données disponibles :

- Spécialité et catégorie du prescripteur et de l'exécutant (médecin généraliste, infirmier, etc.).

Quelles données sont disponibles dans le SNDS ? (2/3)

La base du PMSI (Programme de Médicalisation des Systèmes d'Information) contient l'ensemble des données hospitalières et est géré par l'Agence Technique de l'Information sur l'Hospitalisation (ATIH).

12 millions d'hospitalisations par an

 **Consommation de soins en établissement (hôpitaux, cliniques)**

Données disponibles :

- Actes médicaux pendant le séjour
- Prescriptions de médicaments appartenant à la liste en sus (médicaments innovants et coûteux ainsi que les dispositifs médicaux implantables) - contrairement au DCIR, le PMSI ne contient pas d'informations individuelles sur la prescription des autres médicaments
- Actes et consultations externes (actes et consultations réalisées dans le cadre de l'activité externe de l'établissement de santé, ensemble des prestations hospitalières sans hospitalisation)

 **Informations sur le séjour hospitalier**

Données disponibles :

- Pour l'ensemble des types de séjour (y compris les hospitalisations à domicile), sont recueillies notamment les dates de séjour, les diagnostics principaux et reliés et les niveaux de sévérité.
- D'autres informations sont enregistrées en fonction du type de séjour.

Quelles données sont disponibles dans le SNDS ? (3/3)

Le CépiDc contient l'ensemble des causes médicales de décès et est géré par l'Institut National de la Santé Et de la Recherche Médicale (INSERM).

≈500 000 certificats de décès/an



Causes médicales de décès

Données disponibles :

- Cause initiale (principale) du décès : maladie ou traumatisme étant à l'origine de la séquence des événements morbides ayant entraîné la mort ;
- Cause(s) associée(s) : états morbides, facteurs ou états physiologiques ayant contribué au décès, mais sans être directement à l'origine de la cause initiale
- Date de décès, commune de décès, lieu de décès (domicile, hôpital, voie publique, etc.), informations spécifiques en cas de décès périnatal, etc.



Le SNDS ne contient donc pas que les données associées à des remboursements de l'Assurance Maladie. **Il contient aussi, par exemple, les causes médicales de décès**

Que retenir du SNDS ?



Le SNDS est une constellation de bases pré-existantes alimentées par des flux de données (PMSI, DCIR, CépiDc, SI-DEP, Vaccin COVID...)



Dans le SNDS, ces bases sont déjà reliées entre elles pour multiplier les opportunités de recherche et d'innovation



Mais peut-on relier une base (y-compris hors SNDS) avec des données issues de la base principale du SNDS ? (ex : la base Diane utilisée par le CHU de Nice)

Oui : ces liens sont appelés “appariements”.

Où se trouvent les données utiles pour l'équipe du CHU de Nice ?

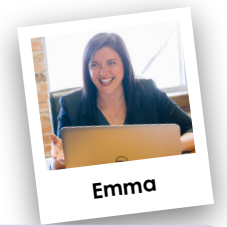


L'équipe du CHU de Nice a besoin de l'historique des actes médicaux de chacun des patients au cours des deux ans avant qu'il intègre la cohorte Diane (c.a.d de 2008 à 2013) pour réaliser son étude.

Dans quelle partie du SNDS (base principale) peut-elle les trouver ?

Comment s'appelle le fait d'utiliser ces données du SNDS pour compléter la base Diane ?

Où se trouvent les données utiles pour l'équipe du CHU de Nice ?



L'équipe du CHU de Nice a besoin de l'historique des actes médicaux de chacun des patients au cours des deux ans avant qu'il intègre la cohorte Diane (c.a.d de 2008 à 2013) pour réaliser son étude.

Dans quelle partie du SNDS (base principale) peut-elle les trouver ?

DCIR

Certains actes médicaux sont réalisés en ville : on n'a pas besoin d'aller à l'hôpital pour faire un examen de biologie par exemple.

PMSI

Néanmoins certaines interventions ne se déroulent qu'à l'hôpital : notamment toutes les hospitalisations et examens complexes.

Pour avoir des informations complètes sur les actes médicaux, il faut des données contenues dans le **PMSI** et dans le **DCIR**

Comment s'appelle le fait d'utiliser ces données du SNDS pour compléter la base Diane ?

Cela s'appelle **apparier** les bases et cela peut se faire de plusieurs façons que nous allons voir dans la prochaine partie.

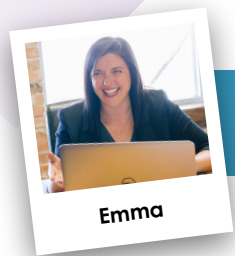


SNDS X

Plan

1. Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane
2. L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?
3. **Comment fait cette équipe en pratique pour compléter la base du projet Diane ?**
4. Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.
5. Et si la base Diane était encore en cours de constitution ? Ce serait différent ?

En pratique, c'est simple un appariement ?



UN 1ER CAS : L'APPARIEMENT DIRECT

L'équipe d'Emma souhaite étudier les actes médicaux réalisés au cours des 2 ans précédant le diagnostic du diabète.

Premier cas :

Les responsables de la mise en oeuvre du projet Diane avaient recueilli le Numéro de Sécurité Sociale des patients (aussi appelé NIR**);*

Certaines informations indispensables à la réalisation de la recherche du CHU de Nice étaient absentes de la base Diane.

Emma et son équipe souhaitent donc compléter la base Diane avec des données issues du SNDS.

* La collecte et le stockage du NIR sont strictement encadrés. En particulier, les NIR des patients ne sont jamais accessibles au promoteur et ils ne peuvent être conservés de façon pérenne : ils sont effacés une fois l'appariement effectué.

** Il est aussi possible de recueillir des informations permettant de reconstruire le NIR.

Comment fonctionne l'appariement direct ?

La base du SNDS et la base Diane peuvent **être reliées directement** puisqu'elles disposent toutes les deux de pseudonymes liés aux NIR des patients* : on parle d'appariement **direct**.

La CNAM est capable de faire le lien entre le NIR et le NIR Haché*

Nir Haché	CP RÉSIDENCE	SEXE	ÂGE	...	CODE ACTE	DATE ACTE
df6063ab	18300	Homme	46		ADQP014	25/12/2012
8d9ea21e	83100	Homme	40	...	ADQP014	15/03/2011
c034b215	75003	Femme	37		HBLD634	16/03/2014



Base SNDS

Nir	ID patient	ID patient	POIDS (kg)	TAILLE (cm)	TYPE DE DIABÈTE
1760483	100	100	65	159	DT1
2841293	101	101	44	150	DT2



Base Diane



Table des NIR séparée

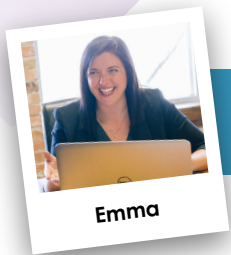
ID patient	CP RÉSIDENCE	SEXE	ÂGE	...	CODE ACTE	DATE ACTE	POIDS (kg)	TAILLE (cm)	TYPE DE DIABÈTE
100	18300	Homme	46		ADQP014	25/12/2012	65	159	DT1
101	75003	Femme	37		HBLD634	16/03/2014	44	150	DT2



Base issue de l'appariement

Note : Cet exemple est purement schématique afin d'illustrer les informations que des bases pourraient avoir en commun. En pratique, la réalisation d'un appariement ne se passera pas aussi directement !**

En pratique, c'est simple un appariement ?



UN 2EME CAS : L'APPARIEMENT INDIRECT

L'équipe d'Emma souhaite étudier les actes médicaux réalisés au cours des 2 ans précédant le diagnostic du diabète.

Deuxième cas :

Les responsables de la mise en oeuvre du projet Diane n'avaient pas recueilli le Numéro de Sécurité Sociale des patients;

Certaines informations indispensables à la réalisation de la recherche du CHU de Nice étaient absentes de la base Diane.

Emma et son équipe souhaitent donc compléter la base Diane avec des données issues du SNDS.

Comment fonctionne l'appariement indirect ?

En l'absence de possibilité d'utiliser le NIR, seul moyen de pouvoir faire le lien avec le pseudonyme SNDS, pour retrouver le même patient dans les deux bases, il est possible de **retrouver les données d'un même patient à partir d'informations communes** (sexe, âge, dates...). De là, si une même combinaison est présente* dans les deux bases on peut réaliser un appariement dit **indirect**.

+

CODE POSTAL RESIDENCE	SEXE	ÂGE	DATE DÉBUT HOSPITALISATION	DATE FIN HOSPITALISATION	CODE ACTE	DATE ACTE
18300	Femme	46	25*/12/2012	02/01/2013	ADQP014	25/12/2012
75003	Femme	37	14/03/2014	18/04/2014	HBLD634	16/03/2014
83100	Homme	40	14/03/2011	21/03/2011	ADQP014	15/03/2011



Base SNDS

SEXE	ÂGE	CODE POSTAL RESIDENCE	POIDS (kg)	TAILLE (cm)	TABAGISME	ALCOOLISME	TYPE DE DIABÈTE	DATE DÉBUT HOSPI	DATE FIN HOSPI
Femme	44	18300	65	159	NON	OUI	DT1	24*/12/2012	02/01/2013
Femme	37	75003	44	105	OUI	NON	DT2	14/03/2014	18/04/2014



Base Diane

=

CODE POSTAL RESIDENCE	SEXE	ÂGE	DATE DÉBUT HOSPI	DATE FIN HOSPI	CODE ACTE	DATE ACTE	POIDS (kg)	TAILLE (cm)	TABAGISME	ALCOOLISME	TYPE DE DIABÈTE
18300	Femme	46	25/12/2012	02/01/2013	ADQP014	25/12/2012	65	159	NON	OUI	DT1
75003	Femme	37	14/03/2014	18/04/2014	HBLD634	16/03/2014	44	105	OUI	NON	DT2



Base issue de l'appariement

Note : Cet exemple est purement schématique afin d'illustrer les informations que des bases pourraient avoir en commun.

Que retenir d'un appariement de données ?

Un appariement de données vise à enrichir une base de données avec des informations complémentaires issues d'une (ou plusieurs) autres bases en les reliant entre elles.

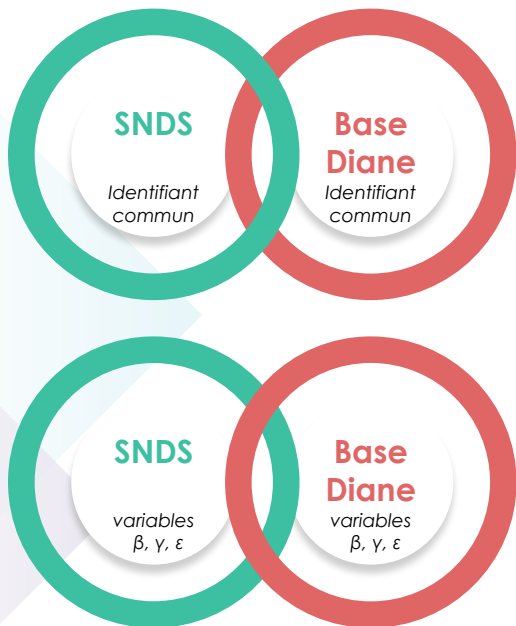
2 types d'appariements existent :

L'appariement DIRECT :

Les données du SNDS et de la base Diane sont reliées directement lorsqu'un patient de la base Diane peut être directement relié à un patient du SNDS grâce au NIR (ou aux éléments permettant de le reconstruire), qui permet de remonter au pseudonyme dans le SNDS (= **numéro de sécurité sociale**).

L'appariement INDIRECT :

La base Diane et le SNDS, sans identifiant unique commun, peuvent être reliés **grâce à un ensemble de variables communes dont la combinaison est suffisamment discriminante pour retrouver toute personne (ou presque) de la base Diane dans le SNDS** (ex : sexe du patient + année de naissance + commune de résidence + date d'hospitalisation).



Plan

1. Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane
2. L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?
3. Comment fait cette équipe en pratique pour compléter la base du projet Diane ?
4. **Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.**
5. Et si la base Diane était encore en cours de constitution ? Ce serait différent ?

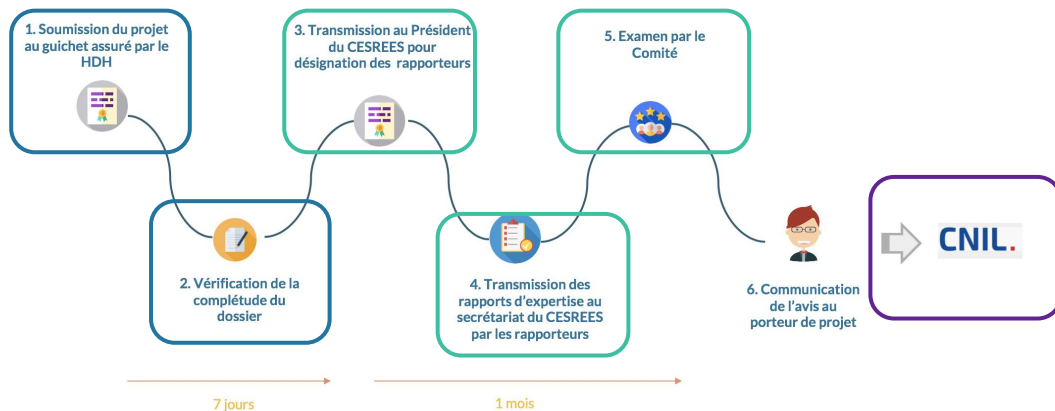
L'équipe du CHU de Nice doit respecter un cadre réglementaire strict

Si on résume la structure de l'étude Diane 2.0 :

- Elle utilise une base de données personnelles de santé déjà existante (la base Diane originale).
- Il n'est pas possible en pratique de réinformer individuellement les patients qui ont participé à Diane.
- Elle veut compléter cette base avec des données issues du SNDS (historique des actes sur 2 ans).

L'étude Diane 2.0 est régie par un cadre réglementaire strict : elle passe par la **procédure standard***. Celle-ci mobilise des instances qui donnent leur accord sur l'intérêt de l'étude et le respect des intérêts des personnes.

Cela impose des étapes de validation **réglementaires, méthodologiques, scientifiques, éthiques** et **technologiques** :



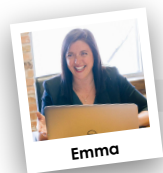
* Il existe d'autres procédures simplifiées d'accès aux données si le projet répond à certains critères. Une description détaillée de tous les accès est disponible [ici](#).

Étape 1 : Dépôt du dossier au HDH - guichet unique



- Le dépôt de la demande d'autorisation s'effectue en ligne auprès du HDH, sur la plateforme [Mon projet](#) (accessible à partir du site internet du HDH). **Le HDH est le guichet unique vers les données du SNDS.**
- Le dossier de demande d'autorisation doit **impérativement** contenir :
 - Un **protocole scientifique** comprenant une **expression de besoins des données du SNDS** précisant le périmètre des données souhaitées.
 - Un **résumé** ;
 - Les **déclarations d'intérêts** des responsables de traitement et de mise en œuvre ;
 - La **demande d'autorisation CNIL** pré-remplie et signée par le responsable de traitement ;
 - La **lettre d'information aux patients** le cas échéant ;
 - La **check-list** synthétisant les points d'attention.

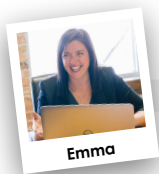
Étape 1 : Dépôt du dossier au HDH - le starter kit



- L'ensemble des modèles ainsi qu'une documentation pédagogique sont mis à la disposition des porteurs de projet dans le [starter kit](#), qui comprend :
 - Des **ébauches de tous les documents nécessaires** pour le dossier dont le protocole scientifique et son résumé, les déclarations d'intérêts des différents partis, de l'expression de besoin du SNDS...
 - Des **guides pédagogiques** sur comment compléter les différents documents et les éventuels documents annexes.
 - Une **checklist** pour s'assurer de ne rien oublier.
- Le HDH vérifie la complétude du dossier et le transmet, le cas échéant, sous 7 jours ouvrés, **au CESREES** pour **évaluation**.



Étape 2 : L'avis du CESREES - validation éthique et scientifique



- Le **CESREES** (ex-CEREES) - **Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé** - est chargé de rendre, avant la saisine de la CNIL, des avis sur les projets d'études nécessitant le recours à des données de santé.
- Le CESREES se prononce sur :
 - La **finalité** et la **méthodologie** de la recherche ;
 - La **nécessité** du recours à des données de santé à caractère personnel ;
 - La pertinence **éthique** ;
 - La qualité **scientifique** du projet ;
 - Le cas échéant, sur le caractère d'**intérêt public** que présente le projet.
- Il est composé d'une **vingtaine de membres nommés par arrêté**. Un **réseau d'experts** extérieurs vient le compléter et son secrétariat est assuré par le HDH.

Étape 2 : L'avis du CESREES - quels projets regarde-t-il ?



Le CESREES examine :

Les projets utilisant des données issues de sources déjà existantes telles que :

- les registres
- les cohortes
- les enquêtes
- les bases médico-administratives
- les dossiers médicaux des établissements de santé

ET

Lorsqu'il est **impossible d'informer individuellement les personnes de l'utilisation de leurs données** (l'analyse de la situation est menée par la CNIL). Par exemple, si on ne dispose d'aucune donnée de contact.



Le CESREES n'examine pas

Les **projets de recherche impliquant la personne humaine (RIPH)** telles que définies dans le code de la santé publique. Un consentement ou une non-opposition (en fonction de la catégorie de la recherche) de doit alors être demandé. Par exemple dans le cadre :

- d'essais cliniques
- de la constitution de biobanques

Ces projets doivent être soumis à un Comité de Protection des Personnes (CPP) avant d'être mis en oeuvre.

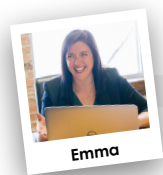
Les **projets conformes à une méthodologie de référence** parmi celles encadrant les projets de recherche non-RIPH*.

Étape 2 : L'avis du CESREES - réception de l'avis



- Le **CESREES** évalue ~ 400 dossiers par an.
Les dates limites de soumission sont précisées pour chaque séance dans le [calendrier du CESREES](#), disponible sur le site internet du HDH. Le comité dispose d'un **délai réglementaire d'1 mois pour se prononcer, renouvelable 1 fois**, à partir de la réception du dossier complet par ce dernier.
- L'avis du CESREES peut être :
 - Favorable ;
 - Favorable avec recommandations ;
 - Réservé (dans ce cas, le dossier peut être présenté au **maximum 3 fois** en comité) ;
 - Défavorable.
- Le **sens de l'avis** est communiqué au porteur de projet 48 heures après la séance du CESREES et l'avis formalisé est transmis au maximum 7 jours après celle-ci.

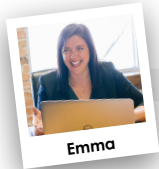
Étape 3 : L'autorisation de la CNIL - instruction de la demande



- Sur demande du porteur de projet, le **HDH transmet le dossier à la CNIL**.
- Les services de la CNIL **instruisent la demande d'autorisation** ; Ils pourront si besoin contacter le porteur de projet, par voie électronique, pour toutes précisions complémentaires.
- La CNIL dispose d'un **délai réglementaire de 2 mois** (à compter de la réception de la demande) **renouvelable une fois** pour transmettre sa décision au porteur de projet.
- Une fois l'autorisation obtenue, la mise à disposition des données sera réalisée par la CNAM (ou le HDH). Cette dernière étape peut prendre quelques mois. Cette étape est décrite plus en détail dans le module sur les accès.*



Étape 3 : L'autorisation de la CNIL - que vérifie-t-elle ?



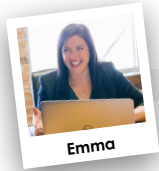
Emma



- Dans le domaine du numérique, la CNIL est le **régulateur des données personnelles**.
- C'est une **autorité indépendante de protection des données personnelles** composée de juristes et d'experts techniques.
- Dans le cadre de ses missions :
 - **veille à la conformité des professionnels** avec la réglementation en vigueur (dont le RGPD et la Loi informatique et libertés)
 - **et aide les particuliers** à maîtriser leurs données personnelles et exercer leurs droits.



Étape 3 : L'autorisation de la CNIL - que doit faire le CHU ?



Le CHU de Nice, *responsable du traitement**, doit être en mesure de démontrer à tout moment sa conformité aux réglementations en traçant toutes les démarches entreprises (principe d'*accountability*).

Il faudra notamment être capable avant, puis tout au long du traitement de :

- **formaliser les rôles et responsabilités du responsable de traitement et des sous-traitants impliqués dans le traitement de données ;**
- **justifier d'une finalité (c'est à dire d'un objectif) d'intérêt public ;**
- **justifier scientifiquement de la nécessité des données traitées** (nature, étendu, profondeur des données) et de l'accès à ces données (pseudonymisées ou directement identifiantes) par les professionnels impliqués dans le projet ;
- **garantir l'information et l'exercice de leurs droits par les personnes ;**
- **préciser la durée de conservation** et l'adapter à la finalité de l'étude ainsi qu'à la nature des données concernées ;
- **garantir la sécurité des données et documenter les actions menées** (hébergement, traçabilité des accès, etc.) ;
- **réaliser une analyse d'impact** sur la protection des données pour le traitement envisagé ;
- **coopérer avec l'autorité de contrôle ;**
- etc.

Plan

1. Une équipe du CHU de Nice souhaite utiliser la base de données de la cohorte Diane
2. L'équipe veut compléter la base du projet Diane avec des données médico-administratives : que peut-elle trouver dans le SNDS pour ce faire ?
3. Comment fait cette équipe en pratique pour compléter la base du projet Diane ?
4. Les données que voudrait utiliser l'équipe du CHU sont très sensibles : il faut respecter un cadre réglementaire strict et rigoureux.
5. **Et si la base Diane était encore en cours de constitution ? Ce serait différent ?**

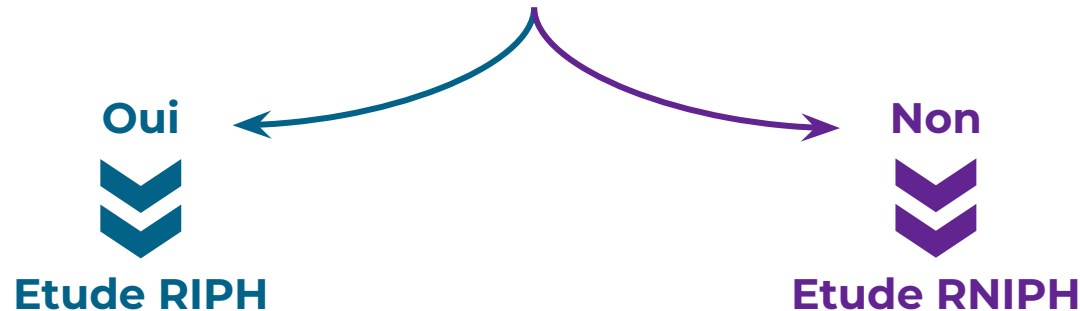
Si la base était en cours de constitution, le cadre réglementaire serait différent...

Que la base Diane 2.0 soit **en cours de constitution signifie que les données sont en train d'être (ou vont être) récoltées** auprès des participants à l'étude. Or, pour répondre à son objectif, l'étude Diane 2.0 doit collecter des informations médicales de façon prospective auprès des patients et de leurs médecins.

L'étude Diane 2.0 implique activement des personnes humaines : elle rentre donc dans la catégorie de la **Recherche Impliquant la Personne Humaine (RIPH)** qui a un parcours de validation différent des **Recherches N'Impliquant pas la Personne Humaine (RNIPH)**.

Il est important de savoir caractériser son projet :

Mon projet est-il une recherche impliquant la personne humaine ou non ?*



La procédure classique est différente pour des études RIPH et RNIPH

La recherche est une RIPH

Jalons clés

Avis du CPP



45 j

Autorisation CNIL



de 2 à 4 mois

Pour en savoir plus sur les modalités de dépôt d'une demande d'accès aux données pour la réalisation d'une RIPH, contactez la [CNRIPH](#) ou consultez [le site du Ministère de la Santé et des Solidarités](#).

La recherche est une RNIPH

Jalons clés

Avis du CESREES



1 mois

Autorisation CNIL



de 2 à 4 mois

Pour en savoir plus sur les modalités de dépôt d'une demande d'accès aux données pour la réalisation d'une RNIPH, consultez le [site du HDH](#).

Qu'avons-nous appris aujourd'hui ?



Un projet d'utilisation de données est régi par un **cadre réglementaire strict qui dépend de l'origine, du moment et du moyen de collecte des données ainsi que de la finalité du projet.**



Les **bases de données déjà existantes** sont **utiles** pour la recherche en santé mais le contenu de chaque base est **limité**.



Il est possible de **compléter une base de données avec les informations contenues dans les bases médico-administratives du SNDS** en procédant à un **appariement**, qui peut être **direct** avec le numéro de sécurité sociale ou bien **indirect** grâce à des données présentes dans les deux bases.



Les projets utilisant des bases de données personnelles de santé déjà existantes sont des études de type RNIPH qui doivent être **validées sur les plans scientifique, méthodologique et éthique par le CESREES*** (comité indépendant qui émet un avis sur les projets) et **en termes de conformité réglementaire par la CNIL** (autorité indépendante de protection des données personnelles) qui est saisie du projet s'il n'est pas conforme à une méthodologie de référence.

Évaluons nos connaissances !

Nous vous proposons de faire un test rapide en 2 questions sur les informations que nous avons vues ensemble aujourd'hui.

Question 1 :

Un porteur de projet dispose d'une base de données récoltées au cours du suivi des patients (base X). Celle-ci est incomplète : il manque certaines prescriptions réalisées en ville.

Le porteur de projet peut-il compléter la base existante avec des données du SNDS, s'il en a besoin ?

- a. Oui, la CNAM peut toujours trouver les patients de la base X dans le SNDS, c'est magique !
- b. Oui, si le porteur de projet a demandé le numéro de sécurité sociale des participants de l'étude.
- c. Oui, si la base X contient assez d'informations en commun avec le SNDS (âge, sexe, dates d'hospitalisation, prescriptions...)
- d. Non, cette pratique est strictement prohibée.

Bonne réponse :

- a. **Oui, la CNAM peut toujours trouver les patients de la base X dans le SNDS, c'est magique !**
- b. **Oui, si le porteur de projet a demandé le numéro de sécurité sociale des participants de l'étude.**
- c. **Oui, si la base X contient assez d'informations en commun avec le SNDS (âge, sexe, dates d'hospitalisation, prescriptions...)**
- d. **Non, cette pratique est strictement prohibée.**

Explication:

Le SNDS peut être utilisé à des fins de recherche **soit en tant que source unique de données** (études épidémiologiques par exemple), soit **en conjonction avec d'autres sources de données** (ici, la base X) dans un processus nommé **appariement**. Ce type de procédé est particulièrement encadré et devra démontrer des mesures de sécurité pointues, vérifiées par la CNIL. Pour ce faire, il faut être capable de retrouver les patients de la base X dans le SNDS. Deux types d'appariement existent :

L'appariement **DIRECT** :

Les données du SNDS et de la base Diane sont reliées directement lorsqu'un patient de la base Diane peut être directement relié à un patient du SNDS grâce au NIR (ou aux éléments permettant de le reconstruire), qui permet de remonter au pseudonyme dans le SNDS (= numéro de sécurité sociale).

L'appariement **INDIRECT** :

La base X et le SNDS, sans identifiant unique commun, peuvent être reliés grâce à un ensemble de variables communes dont la combinaison est suffisamment discriminante pour retrouver toute personne (ou presque) de la base X dans le SNDS (ex : sexe du patient + année de naissance + commune de résidence + date d'hospitalisation).

Question 2 :

Une équipe dispose d'une base de données issues de défibrillateurs implantés autonomes (c'est comme des pacemakers connectés) et voudrait les compléter avec des données du SNDS pour développer un algorithme d'intelligence artificielle permettant le suivi amélioré des patients atteints de troubles cardiaques.

Que doit faire le porteur de ce projet pour compléter la base existante avec des données du SNDS ?

- a. Il doit justifier scientifiquement et éthiquement la nécessité de l'appariement.
- b. Il doit définir au préalable quelles données du SNDS sont nécessaires pour leur étude.
- c. Il peut demander toutes les données contenues dans le SNDS pour être sûr de ne rien rater.
- d. Il peut demander directement à la CNAM un accès aux données nécessaires.
- e. Il passe par le Health Data Hub pour soumettre une demande adaptée qui sera ensuite analysée par les organismes compétents.
- f. Il doit s'assurer d'avoir les données nécessaires pour pouvoir réaliser un appariement.

Bonne réponse :

- a. Il doit justifier scientifiquement et éthiquement la nécessité de l'appariement.
- b. Il doit définir au préalable quelles données du SNDS sont nécessaires pour leur étude.
- c. Il peut demander toutes les données contenues dans le SNDS pour être sûr de ne rien rater.
- d. Il peut demander directement à la CNAM un accès aux données nécessaires.
- e. Il passe par le Health Data Hub pour soumettre une demande adaptée qui sera ensuite analysée par les organismes compétents.
- f. Il doit s'assurer d'avoir les données nécessaires pour pouvoir réaliser un appariement.

Explication:

En fait, ce projet existe : c'est le projet HYDRO, un des lauréats des appels à projet du Health Data Hub qu'on aborde dans cette même formation !

Ce projet souhaite **utiliser les données du SNDS pour développer un algorithme d'intelligence artificielle** pour améliorer le suivi des patients atteints de troubles cardiaques. Pour ce faire, il évolue au sein d'un cadre strict grâce à l'aide du HDH. En montant le dossier de l'étude, le porteur du projet a **justifié le besoin des données et a expliqué la méthodologie envisagée**.

Celui-ci est transmis, par le HDH aux autres instances compétentes :

- un comité éthique et scientifique indépendant qui fournit un avis sur le projet
- la CNIL (autorité indépendante de protection des données personnelles) qui, entre autres, en évalue la conformité réglementaire

Pour plus d'informations à ce sujet et sur les projets lauréats des appels à projets, rdv sur le site du HDH.



document version : 1.3 - 10 Mars 2023.

conception et expertise :

Sanoia
Real World Digital CRO

photos : unsplash.com ; icones : thenounproject.com (licence)