GUIDE PRATIQUE DE REFERENCE POUR L'APPARIEMENT DES DONNEES DE L'EDS DE L'AP-HP AVEC LE SNDS

[Comment apparier les données de l'EDS AP-HP avec les données du SNDS ?]

Novembre 2025

eds AP-HP Libérons le potentiel de nos données de santé

C1 - Interne

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

SOMMAIRE

GLOSSAIRE 4

INTRODUCTION 5

L'APPARIEMENT AUX DONNEES DU SNDS 6

ROLE ET RESPONSABILITES DE CHAQUE PARTIE PRENANTE 8

LES ETAPES DU PROCESSUS D'APPARIEMENT 10

Etape 1 : Prendre contact avec l'URC 13

Etape 2 : Etudier la faisabilité du projet avec les équipes de l'EDS 15

Etape 3 : Accéder aux données de l'EDS de l'AP-HP 15

3.1 Procédure avec passage devant le CSE 15

3.1.1 Examen du dossier par le CSE 15

3.1.2 Partenariat externe 15

3.1.3 Requête de sélection des patients 15

3.1.4 Livraison des données 16

3.2 Procédure sans passage devant le CSE 16

Etape 4 : Valider le projet devant le CESREES et la CNIL 17

4.2 La constitution du dossier réglementaire 17

4.2 L'examen du dossier par le CESREES 18

4.3 L'examen du dossier par la CNIL 19

Etape 5 : Signer la convention 20

Etape 6 : Suivre la formation SNDS 20

Etape 7 : Réaliser techniquement l'appariement..... 21

7.1 Préparation des données nécessaires de l'EDS sur l'espace de travail de l'AP-HP22

7.2 Transmission des données sur la plateforme d'appariement..... 23

7.3 Appariement des données SNDS avec les données de la table d'appariement23

7.4 Transfert des données médicales sur la plateforme de traitement 25

NOTRE OFFRE D'ACCOMPAGNEMENT INNOVATION ET DONNEES.....26

ANNEXE 27

Projets passés de chaînage EDS / SNDS 27

Les trois circuits d'appariement possible 27

Appariement direct via le NIR 27

Appariement avec reconstitution du NIR..... 28

Appariement indirect..... 28

Liste des actions à réaliser pour le chaînage des données avec le SNDS..... 29

GLOSSAIRE

ALD	Affection de Longue Durée
AP-HP	Assistance Publique – Hôpitaux de Paris
CASD	Centre d'Accès aux Données de Santé
CCAM	Classification Commune des Actes Médicaux
CESREES	Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé
CIM-10	Classification Internationale des Maladies - 10e révision
CNAM	Caisse nationale d'Assurance maladie
CNAV	Caisse Nationale d'Assurance Vieillesse
CNIL	Commission nationale de l'informatique et des libertés
CSE	Comité Scientifique et Ethique
CSSI	Consultation Système d'Information
DPI	Dossier Patient Informatisé
DPO	Data Protection Officer
DRCI	Direction de la Recherche Clinique et de l'Innovation
DSN	Direction des Services Numériques
EDS	Entrepôt de Données de Santé
FINESS	Fichier National des Établissements Sanitaires et Sociaux
GHM/GHS	Groupe Homogène de Malades / Groupe Homogène de Séjours
HDH	Health Data Hub
HTTPS	HyperText Transfer Protocol Secure
ISD	Ingénierie et Science des Données
MR004	Méthodologie de Référence
NIR	Numéro d'Inscription au Répertoire
NLP	Natural Language Processing
PIA	Privacy Impact Assessment
PMSI	Programme de Médicalisation des Systèmes d'Information
RGPD	Règlement Général sur la Protection des Données
RSSI	Responsable de la Sécurité des Systèmes d'Information
RUPP	Relation Utilisateurs et Pilotage Projets
SAR	Subject Access Request
SNDS	Système National des Données de Santé
SNGI	Système National de Gestion des Identifiants
SNIIRAM	Système national d'information inter-régimes de l'Assurance maladie
URC	Unité de Recherche Clinique

INTRODUCTION

L'appariement entre les données de l'Entrepôt de données de santé (EDS) de l'AP-HP et celles du Système national des données de santé (SNDS) représente une opportunité pour enrichir les projets de recherche en santé publique, en épidémiologie, en médecine ou encore en intelligence artificielle appliquée à la santé. L'EDS offre un niveau de détail important sur l'état de santé des patients suivis dans les 38 hôpitaux de l'AP-HP, incluant les comptes-rendus médicaux, résultats biologiques, prescriptions, imagerie, etc. De l'autre côté, le SNDS rassemble l'ensemble des données médico-administratives de l'assurance maladie : actes, hospitalisations, consultations, médicaments remboursés, Affection de longue durée (ALD), causes médicales de décès, couvrant l'ensemble de la population nationale.

Ainsi, en croisant ces deux sources de données, les chercheurs peuvent combiner des données cliniques fines issues des établissements hospitaliers de l'AP-HP avec des données médico-administratives exhaustives couvrant l'ensemble du parcours de soins des patients au niveau national.

Toutefois, cette démarche s'inscrit dans un cadre réglementaire, méthodologique et technique qui nécessite l'implication de nombreux institutionnels : AP-HP, Caisse nationale d'Assurance maladie (CNAM), Commission nationale de l'informatique et des libertés (CNIL), Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) et Health Data Hub (HDH) ainsi que le respect de procédures précises. Obtenir l'autorisation d'appariement et pouvoir accéder aux données demandent une préparation rigoureuse, une anticipation des délais et un strict respect des exigences en matière de protection des données personnelles.

C'est pour répondre à ces besoins que ce guide a été conçu. Il vise à fournir aux investigateurs un accompagnement pas-à-pas, depuis le montage du projet de recherche jusqu'à l'exploitation des données appariées.

En s'appuyant sur des retours d'expérience concrets de projets déjà accompagnés par les équipes de l'EDS, ce guide décrit la procédure pour apparier ces deux bases et propose aussi des conseils pratiques et des outils pour limiter les risques, les délais d'instruction et pour garantir la faisabilité de votre projet de recherche.

L'APPARIEMENT AUX DONNEES DU SNDS

L'EDS de l'AP-HP et le SNDS sont deux sources de données de santé aux caractéristiques différentes. Le SNDS[1] rassemble l'information sur l'ensemble des prestations de soins remboursées par l'assurance maladie que ces prestations soient fournies en ville via le Système national d'information inter-régimes de l'Assurance maladie (SNIIRAM) ou à l'hôpital via le Programme de médicalisation des systèmes d'information (PMSI). Ce système d'information a été mis en place pour assurer la facturation et le remboursement des soins : l'information est donc quasi-exhaustive sur le territoire national mais se limite à des informations nécessaires à la facturation et au remboursement ainsi qu'aux causes de décès. L'EDS de l'AP-HP[2] est une base de données issues du soin qui décrit plus précisément l'état de santé et les actes/soins réalisés (biologie, médicament, imagerie, résultats d'analyse, etc.) mais uniquement pour les patients des hôpitaux de l'AP-HP.

L'appariement entre l'EDS de l'AP-HP et le SNDS consiste à rapprocher les données relatives aux mêmes patients présents dans les deux bases afin de profiter des avantages de chacune d'elle. À titre d'exemple, cela permet de réaliser un phénotypage des patients à partir de critères plus précis que la CIM-10 comme une analyse des comptes rendus ou le résultat d'examen spécifique (biologies, imagerie…) et d'obtenir l'intégralité du parcours de soin des patients comme les consultations en médecine libérale, les traitements pris en ville et le suivi dans d'autres hôpitaux. En épidémiologie, le SNDS pourrait être utilisé pour compter la prévalence d'une pathologie (CIM-10, médicament ou acte spécifique) à l'échelle nationale et les données spécifiques de l'AP-HP permettraient de mieux caractériser les patients sur un sous ensemble de la population en enrichissant les variables du SNDS. Il reste que si le potentiel d'exploitation que représente l'appariement de ces deux bases est évident, la procédure pour y parvenir est longue et couteuse[3]. Il convient donc de s'assurer de la nécessité de cet appariement pour mener à bien la recherche.

[1] Lien vers le site du HDH pour présentation du SNDS.
[2] Consulter le site internet de l'EDS pour plus d'informations.
[3] Le rapport « Fédérer les acteurs de l'écosystème pour libérer l'utilisation secondaire des données de santé » (décembre 2023) recommande de simplifier ces démarches en réduisant notamment la logique d'autorisation systématique.

Sur le plan technique, plusieurs types d'appariement sont possibles :

> L'appariement direct utilise le numéro d'inscription au répertoire (NIR) pour identifier les patients dans les deux bases. Cet appariement est rapide et sûr du point de vue technique mais plus complexe du point de vue réglementaire car les traitements basés sur le NIR, comme ceux basés sur des données nominatives, sont soumis à une réglementation supplémentaire[4].

> L'appariement direct avec reconstitution du NIR consiste à utiliser les données d'état civil (nom, prénom, sexe, date et lieu de naissance) pour retrouver le NIR via le système national de gestion des identifiants (SNGI), géré par la CNAV. Ce procédé permet d'effectuer un chaînage direct à partir de l'état civil. Le HDH prévoit de proposer cette reconstitution dans le cadre de l'offre de service de son concentrateur, mais cela n'est pas possible à ce jour.

> L'appariement indirect s'appuie sur un ensemble de variables communes entre l'EDS et le SNDS. Il s'agit généralement des données du PMSI comme la date de naissance, le genre, l'établissement fréquenté, la date de la visite, la liste des diagnostics CIM-10, etc. Ce type d'appariement est indirect car il repose sur le fait qu'avec suffisamment d'informations (dates de plusieurs visites, liste complètes des diagnostics, etc.), les patients peuvent être identifiés de façon unique dans les bases ce qui permet de rapprocher les identités.

[4] Voir le guide de la CNIL

ROLE ET RESPONSABILITES DE CHAQUE PARTIE PRENANTE

Cinq acteurs interviennent dans le processus de chaînage des données avec le SNDS :

ASSISTANCE PUBLIQUE HÔPITAUX DE PARIS

Les équipes de la Direction de la Recherche Clinique et de l'Innovation (DRCI) et du pôle Innovation et Données de la Direction des Services Numériques accompagnent les investigateurs de l'AP-HP dans les différentes étapes d'élaboration, de réalisation et de valorisation de leurs projets de recherche :

- Mobilisation des expertises médicales pour accompagner la conception du projet de recherche et sa réalisation
- Apport d'expertises techniques complémentaires, notamment en matière de sécurité des systèmes d'information
- Accompagnement des démarches de conformité par rapport à la politique RGPD
- Soutien contractuel dans le cadre de collaborations et partenariats de recherche
- Implication des méthodologistes dans le montage des projets de recherche
- Appui à la valorisation des projets (publications, brevets et communication)
- Implication des méthodologistes dans le montage des projets de recherche et l'analyse des données en lien avec les datascientists et les biostatisticiens

CESREES
Comité Éthique et Scientifique pour les Recherches, les Études et les Évaluations en Santé

Le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) est chargé d'évaluer les projets de recherche 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

L'avis du CESREES est obligatoire pour obtenir une autorisation de chaînage avec le SNDS.

La commission nationale de l'informatique et des libertés (CNIL) est une autorité administrative indépendante en charge de la protection des données personnelles. La CNIL a la charge de veiller à la protection de toute information se rapportant à une personne physique. Dans ce cadre, elle a une mission de conseil, et dispose également d'un pouvoir de contrôle et de sanction.

La CNIL est l'institution qui délivre les autorisations de chaînage attestant de la conformité du projet de chainage à la réglementation relative à la protection des données personnelles (RGPD, loi informatique et libertés, etc.). L'autorisation de la CNIL est nécessaire pour réaliser un chaînage.

Le Health Data Hub (HDH) est le guichet d'entrée unique pour les projets nécessitant une autorisation de chaînage. Le dépôt d'un dossier complet, incluant plusieurs pièces justificatives et la mobilisation d'acteurs clés (DPO, RSSI, etc.), est indispensable pour son instruction par le CESREES puis la CNIL.

Le SNDS, piloté par la caisse nationale de l'assurance maladie (CNAM), a été conçu pour faciliter l'accès et l'ouverture des données de santé qu'elle centralise et met à disposition.

La CNAM joue un rôle central dans ce dispositif : elle est responsable de l'ouverture des données issues du SNDS et accompagne les utilisateurs à chaque étape. Cet accompagnement prend la forme de formations, d'un support dédié et d'une documentation complète.

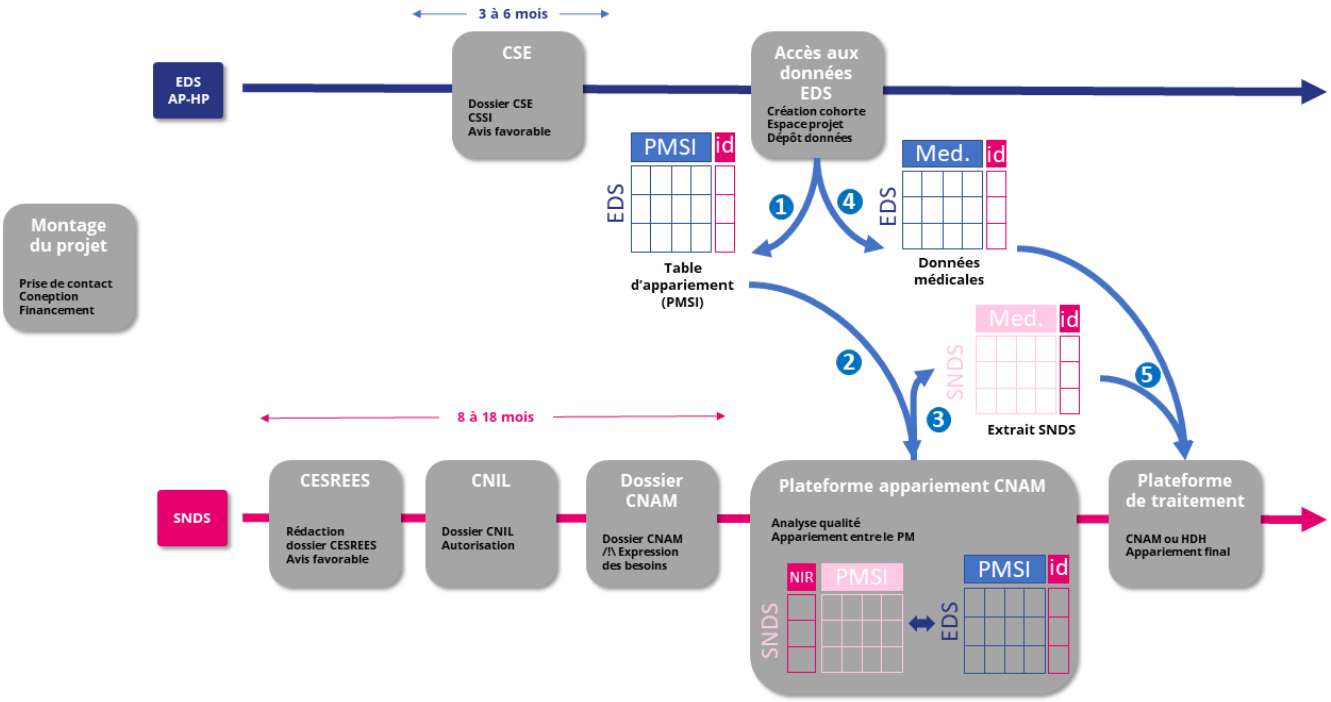
Dans le cadre du chaînage des données, la CNAM intervient à plusieurs niveaux : elle forme les investigateurs de projets, fournit les données nécessaires et assure la contractualisation indispensable à la finalisation de l'appariement.

LES ETAPES DU PROCESSUS D'APPARIEMENT

De manière générale, le parcours se décompose en deux volets complémentaires :

> La demande classique d'accès aux données de l'EDS de l'AP-HP précisant la volonté de faire un chaînage indirect avec les données du SNDS

> La demande d'autorisation auprès de la CNIL et l'accès aux données du SNDS



Plusieurs scenarios sont envisageables pour mener à bien un appariement entre l'EDS et le SNDS. Le scenario principal décrit brièvement ci-dessous est recommandé par défaut, chacune des étapes et des alternatives envisageables font l'objet de descriptions plus approfondies dans les parties suivantes.

1. Prendre contact avec l'URC

Tout projet de recherche doit débuter par un contact avec le coordinateur ou le datascientist de l'URC, qui gère l'accompagnement méthodologique, réglementaire et l'enregistrement de la demande dans l'outil Timetonic.

A noter qu'étant donné la complexité et la durée de la démarche, il est essentiel de disposer d'un financement pour mener à bien ce type de projet de recherche.

2. Etudier la faisabilité du projet (CSSI)

Cette étude de faisabilité, qui prépare le passage devant le CSE de l'AP-HP, permet d'aborder les principales difficultés organisationnelles (ressources, calendrier, réglementaires, etc.) et techniques (accès aux données, appariement, etc.) du projet. Les équipes de l'URC et de l'EDS vous accompagnent dans cette étape.

3. Accéder aux données de l'EDS de l'AP-HP

L'accès à ces données se fait selon la procédure classique qui implique de présenter le projet devant le CSE de l'AP-HP en mentionnant l'intention de chaînage avec le SNDS. Il est important de réaliser cette étape rapidement afin d'éviter d'avoir à attendre la validation par les comités nationaux (CESREES et CNIL) dont les délais sont longs. Un accès rapide aux données permet de commencer le travail d'exploitation des données de l'EDS dont une partie importante est réalisée avant l'appariement avec le SNDS (préparation des tables d'appariement, le cas échéant extraction de données par NLP, réalisation des traitements complexes, etc.). Il est important de noter que les documents cliniques ne quittent pas les espaces Jupyter de l'AP-HP et que tous les traitements NLP doivent donc se faire dans ces espaces sécurisés. Dans certains cas, cette première exploitation peut même faire l'objet d'une ou plusieurs publications scientifiques.

4. Valider le projet devant le CESREES et la CNIL

Cette étape est nécessaire pour accéder aux données du SNDS et pour obtenir l'autorisation d'apparier les deux sources de données. Afin d'éviter tout risque de rejet, le dossier doit être rédigé avec clarté et précision. C'est l'étape la plus longue de la procédure et il faut compter généralement au moins 8 mois pour la réaliser. L'autorisation de la CNIL, transmise directement à la CNAM, permet au projet d'accéder aux données et autorise les personnes qui vont techniquement réaliser les traitements à s'inscrire à la formation obligatoire du SNDS.

5. Suivre la formation SNDS

Une formation, dispensée par la CNAM pour une durée de 4 jours est obligatoire pour accéder aux données du SNDS. Cette formation est valide pendant toute la durée de l'habilitation de l'utilisateur, il n'est donc pas nécessaire de la suivre pour chaque nouveau projet.

6. Signer la convention

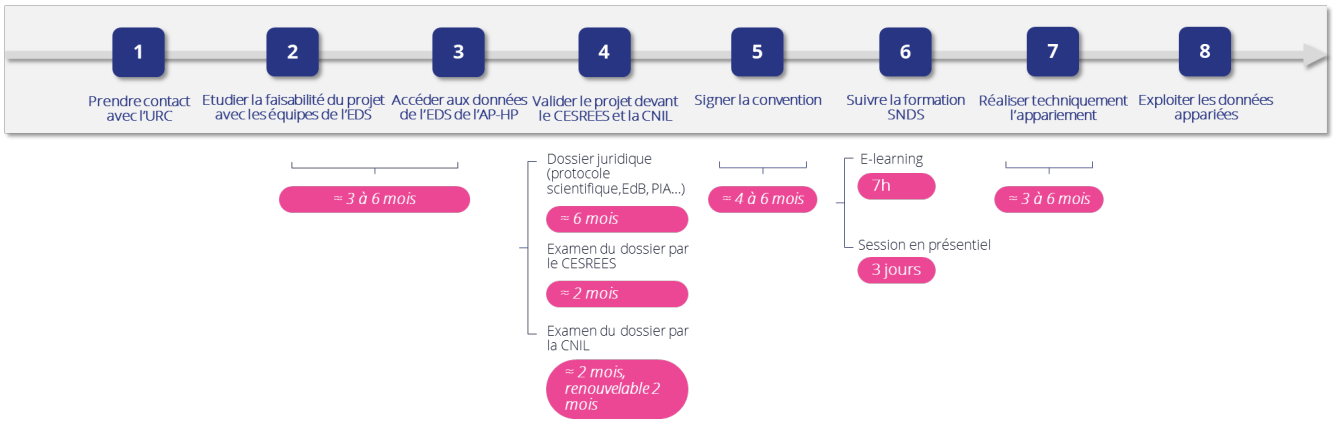
Une convention est établie entre la CNAM et le responsable de traitement afin de définir les modalités de mises à disposition des données dans le périmètre autorisé. La phase de conventionnement, dont la durée varie généralement entre 4 et 6 mois, dépend de la complexité du projet ainsi que du traitement des données demandées.

7. Réaliser techniquement l'appariement

A ce jour, seul l'appariement indirect est possible avec les données de l'EDS de l'AP-HP. Pour le réaliser, il faut extraire de l'EDS les données d'appariement qui seront rapprochées de celles du SNDS. Le SNDS contenant les données du PMSI, ce sont ces données qui constituent l'information la plus fiable pour rapprocher les deux bases. La table d'appariement est donc constituée d'un sous-ensemble des données du PMSI et d'un identifiant unique par patient spécifique pour le projet. Cette table est transmise à la CNAM qui réalise l'appariement avec la version nationale du PMSI en demandant, si nécessaire, des données complémentaires. Par la suite, la CNAM extrait du SNDS les données demandées et les met à disposition du projet dans leur espace de traitement en y accolant l'identifiant unique par patient spécifique pour le projet. Une fois ces données déposées, le reste des données de l'EDS sont transmises sur la même plateforme de traitement avec le même identifiant unique spécifique au projet.

8. Exploiter les données appariées

L'exploitation des données appariées se fait sur la plateforme de la CNAM[5] qui dispose de peu de puissance de calcul et d'une gamme d'outils limitée. Il est donc important d'avoir réalisé, en amont, le plus de traitements possibles, notamment ceux nécessitant des méthodes d'apprentissage statistique.



[5] D'autres choix de plateformes d'appariement sont possible (HDH ou CASD) comme cela est détaillé dans l'étape 7 : Réaliser techniquement l'appariement

La suite de ce document décrit plus en détail chacune des étapes ainsi que les choix possibles pour les investigateurs.

Etape 1 : Prendre contact avec l'URC

Avant d'initier un projet de recherche utilisant les données de l'EDS, il est essentiel de prendre contact avec le coordinateur EDS de votre groupement hospitalo-universitaire (GHU) ou le datascientist de votre URC. Ils vous accompagnent dans les choix méthodologiques, la conception du projet et les premières démarches technico-réglementaires dont la préparation du kit de demande d'accès aux données de l'EDS ainsi que préparation du dossier juridique (demande d'avis auprès du CESREES et autorisation de la CNIL).

Le kit de demande est constitué de six documents :

- La Déclaration Publique d'Intérêt (DPI) : La déclaration publique d'intérêts (DPI) permet d'identifier les éventuels liens d'intérêts du déclarant avec des organismes publics ou privés intervenant dans le champ de compétences de l'AP-HP. Elle mentionne les liens d'intérêts de toute nature, directs ou par personne interposée, que le déclarant a, ou qu'il a eus pendant les cinq années précédant sa prise de fonctions, avec des entreprises, des établissements ou des organismes dont les activités, les techniques et les produits entrent dans le champ de compétence de l'autorité sanitaire au sein de laquelle il exerce ses fonctions ou de l'organe consultatif dont il est membre ainsi qu'avec les sociétés ou organismes de conseil intervenant dans les mêmes secteurs. Elle est rendue publique, y compris en ce qui concerne les rémunérations reçues par le déclarant de la part d'entreprises, d'établissements ou d'organismes mentionnés au troisième alinéa ainsi que les participations financières qu'il y détient. Elle est actualisée à l'initiative de l'intéressé.
- Le formulaire de demande d'accès aux données de l'EDS : Il s'agit d'un formulaire interne permettant de recenser les informations générales du projet, les données nécessaires pour la recherche, les services complémentaires demandés, les financement et partenariats.
- La matrice des habilitations : Il s'agit d'un document permettant de recenser le contact de toutes les personnes qui seront habilitées à accéder à l'espace de travail contenant les données du projet de recherche (nom, prénom, adresse mail, hôpital, fonction…).
- La check-list de conformité à la MR004[6] : La méthodologie de référence MR-004 encadre les traitements de données à caractère personnel à des fins d'étude, évaluation ou recherche n'impliquant pas la personne humaine. Il s'agit plus précisément des études ne répondant pas à la définition d'une recherche impliquant la personne humaine, en particulier les études portant sur la réutilisation de données. La recherche doit présenter un caractère d'intérêt public. Le responsable de traitement s'engage à ne collecter que les données strictement nécessaires et pertinentes au regard des objectifs de la recherche.
- Le Privacy Impact Assessment (PIA) : Le PIA est un outil important pour la responsabilisation des organismes. Elle les aide non seulement à construire des traitements de données respectueux

[6] https://www.cnil.fr/fr/declaration/methodologie-de-reference-04-recherches-nimpliquant-pas-la-personne-humaine-etudes-et-evaluations-dans-le-domaine-de-la-sante

de la vie privée, mais aussi à démontrer leur conformité au Règlement général sur la protection des données (RGPD). Elle est obligatoire pour les traitements susceptibles d'engendrer des risques élevés.

> Le protocole de recherche : Il s'agit d'un document qui décrit de manière détaillée la façon dont le projet de recherche va être mené.

Une fois ce kit constitué, l'URC enregistre la demande dans l'outil Timetonic à l'aide du formulaire de demande en ligne et réalise l'étude, la construction méthodologique et scientifique de votre projet.

Le dossier suit ensuite un circuit de validation spécifique :

> Le PIA est transmis au Délégué à la Protection des Données (DPO) central de l'AP-HP puis au Responsable de la Sécurité des Systèmes d'Information (RSSI) pour approbation qui est garant du respect de la politique générale de sécurité de l'information de l'AP-HP.

> La MR004 est, quant à elle, adressée au DPO pour validation.

> La DPI ainsi que le protocole de recherche sont soumis à l'examen du Comité Scientifique et Ethique (CSE).

NOS CONSEILS / POINTS D'ATTENTION

> Consultez la liste des coordinateurs EDS et datascientists de votre GH ainsi que le kit de demande

> Le processus de chaînage des données de l'EDS avec celles du SNDS peut se révéler long c'est pourquoi il est important d'anticiper dès le départ la procédure et de structurer les différentes étapes avec l'URC afin de gagner en efficacité.

« Pour le chaînage, on n'avait pas anticipé que la procédure prendrait autant de temps. On était partis sur un financement de deux ans, donc heureusement que l'AAP DATAE nous a permis de le prolonger » - Investigateur

Etape 2 : Etudier la faisabilité du projet avec les équipes de l'EDS

En lien avec votre URC, vous devrez saisir l'équipe de l'EDS qui réalisera une étude technique complémentaire. Cette étude, qui prépare le passage devant le Comité scientifique et éthique de l'AP-HP, permet d'aborder les principales difficultés organisationnelles (ressources, calendrier, réglementaires, etc…) et techniques (accès aux données, appariement, etc…) du projet. C'est une étape indispensable pour anticiper les difficultés, planifier les travaux à réaliser et connaître les coûts associés à votre projet.

A l'issue de la réunion de faisabilité (CSSI) organisée entre vous et l'équipe de l'EDS, une étude de faisabilité vous est transmise avec un chiffrage des coûts de mise à disposition des données et de la création de la table d'appariement. Cela vous permettra, le cas échéant, de solliciter un financement adapté aux besoins de votre projet de recherche.

NOS CONSEILS / POINTS D'ATTENTION

Contactez au plus tôt l'équipe support de l'EDS pour planifier une réunion de faisabilité (CSSI) et ainsi vous pourrez anticiper sur votre recherche de financement : id.recherche.support.dsn@aphp.fr

Etape 3 : Accéder aux données de l'EDS de l'AP-HP

3.1 Procédure avec passage devant le CSE

3.1.1 Examen du dossier par le CSE

Sous réserve de votre engagement à couvrir les coûts de mise à disposition des données, l'accès aux données de l'EDS se fait selon la procédure classique qui implique de présenter le projet devant le CSE de l'AP-HP en mentionnant l'intention de chaînage avec le SNDS.

Le CSE étudie votre projet et revient vers vous avec son avis. Si l'avis est favorable, vous serez autorisé à démarrer votre projet.

3.1.2 Partenariat externe

En cas de partenariat externe, des discussions contractuelles seront engagées par la DRCI.

3.1.3 Requête de sélection des patients

Votre URC vous aide à cibler les données nécessaires à votre projet de recherche et à construire la requête définitive, le document de traçabilité et la matrice d'habilitation.

3.1.4 Livraison des données

L'équipe EDS crée un **espace de travail sécurisé** contenant les données pseudonymisées de votre projet et les services numériques demandés.

3.2 Procédure sans passage devant le CSE

Dans le cas du chaînage avec le SNDS, **le passage devant le CSE de l'AP-HP n'est pas obligatoire**. En effet, **l'autorisation du projet par le CESREES et la CNIL prévaut sur l'avis du CSE**. Une alternative à la procédure proposée ci-dessus est donc d'entamer le projet par l'étape 4 (Valider le projet devant le CESREES et la CNIL) ci-dessous puis d'informer le CSE, une fois l'autorisation de la CNIL obtenue, du droit d'accès aux données.

L'avantage de cette alternative est qu'elle limite les démarches administratives puisqu'elle ne nécessite pas la rédaction d'un dossier CSE. Elle sera adaptée aux projets disposant de peu de ressources de chefferie de projet, n'ayant pas encore de ressources à disposition pour traiter les données et n'ayant pas de contraintes temporelles fortes. En suivant cette alternative, l'étape 1 (Prendre contact avec l'URC) et l'étape 2 (Etudier la faisabilité du projet avec les équipes de l'EDS) peuvent être réalisées **ultérieurement** et l'accompagnement de projet peut même être réalisé par le **HDH** qui propose ce type de service. Il faudra toutefois veiller à **réaliser une CSSI avant la transmission du dossier au CESREES afin d'assurer la faisabilité technique du projet** et notamment la disponibilité de l'intégralité des données demandées.

Ce guide fait le choix de ne pas recommander cette alternative, bien qu'elle soit tout-à-fait possible, et **privilégie le passage devant le CSE** pour des raisons opérationnelles. En effet, puisque la procédure d'accès à l'EDS de l'AP-HP est **plus courte**, la procédure avec passage devant le CSE permet **d'accéder rapidement aux données de l'EDS** et ainsi de **commencer la préparation des tables d'appariement**, **l'extraction des variables nécessaires** au projet et, dans certains cas, de **réaliser une première production scientifique uniquement sur les données de l'EDS en parallèle de la procédure réglementaire CESREES/CNIL**. Ce choix permet ainsi de **réduire le temps total d'accès aux données chaînées** puisque les données sont **prêtes à être transmises dès la signature de la convention avec la CNAM**. Ce choix est particulièrement adapté si la personne qui traite les données est disponible dès le début du projet.

Etape 4 : Valider le projet devant le CESREES et la CNIL

La validation du projet devant le CESREES et la CNIL repose sur trois grandes étapes, longues et coûteuses, mais indispensables pour accéder aux données du SNDS et obtenir l'autorisation d'apparier les deux sources de données. Ce sont les étapes les plus longues de la procédure nécessitant généralement un délai d'au moins 8 mois pour les mener à bien (chaque comité dispose de 1 à 2 mois renouvelables pour traiter le dossier et des retards sont possibles).

4.2 La constitution du dossier réglementaire

Délai estimé de 4 mois

Tout projet de chaînage nécessite de compléter le starter kit mis à disposition par le HDH.

Pour déposer une demande d'autorisation, plusieurs documents sont requis :

- Un protocole scientifique : ne seront recevables que les protocoles respectant l'ensemble des consignes de remplissage, et comprenant la totalité des informations demandées dans chaque partie
- Un résumé type
- La checklist des éléments clés du protocole
- L'expression de besoin en données SNDS à joindre en annexe du protocole scientifique avec son guide de remplissage pour vous aider
- La déclaration publique d'intérêt du responsable de traitement et du responsable de mise en œuvre
- La déclaration publique d'intérêt du laboratoire de recherche ou du bureau d'études
- Le formulaire de demande d'autorisation de la CNIL
- L'annexe sur le transfert de données hors de l'Union Européenne le cas échéant
- La lettre d'information aux personnes (le cas échéant) et son guide pédagogique sur la note d'information aux personnes pour vous aider

En complément, d'autres pièces peuvent vous être demandées :

- L'avis du ou des comités scientifiques et/ou éthiques qui ont pu évaluer le projet et la composition de ce(s) comité(s), le cas échéant
- La liste des financeurs de l'étude

Une fois le dossier constitué et complet, il doit être déposé sur la plateforme de dépôt du HDH, avant d'être instruit par le CESREES. L'avis du CESREES est un prérequis avant l'autorisation CNIL.

NOS CONSEILS / POINTS D'ATTENTION

> Il est primordial de rédiger **l'expression de besoin** avec une extrême **rigueur** et de veiller à ce qu'elle soit parfaitement **conforme** à **l'autorisation de la CNIL**. **Toute erreur ou imprécision dans cette documentation pourrait obliger à redemander l'accord de la CNIL**, prolongeant ainsi le processus préliminaire à l'appariement.

> Les **contraintes d'appariement** doivent être anticipées en choisissant **dès le départ** la **méthode** la plus adaptée (direct ou indirect).

> Les **plateformes d'appariement et de traitement des données** (CNAM ou HDH) doivent également être **choisis dès cette étape**.

> Le HDH peut relire l'expression de besoin avant la transmission au CESREES pour éviter les principaux écueils.

4.2 L'examen du dossier par le CESREES

Délai estimé de 2 mois (un mois renouvelable une fois)

Une fois le dossier déposé sur la plateforme de dépôt du HDH, **il suit plusieurs étapes d'examen avant d'obtenir l'autorisation finale de la CNIL.**

La première étape consiste en l'examen du dossier par le CESREES. Ce comité évalue la rigueur méthodologique du projet, sa pertinence scientifique, ainsi que la nécessité du recours à des données à caractère personnel. Il s'assure également que les données demandées sont cohérentes avec la finalité du traitement envisagé.

Au cours de cet examen, **le CESREES peut émettre des demandes de compléments ou d'ajustements**. L'investigateur est alors invité à y répondre dans les meilleurs délais afin de permettre la poursuite de l'instruction. **Ces échanges sont fréquents et visent à clarifier certains points du dossier ou à garantir le respect des exigences éthiques et scientifiques.** Une expression des besoins claire et exhaustive permet de **limiter ces allers-retours chronophages qui vont ralentir l'avancement du projet.**

Les pouvoirs du CESREES sont définis par l'Arrêté du 5 mai 2021 : « *à l'expiration d'un délai d'un mois à compter de la date réception du dossier complet par le comité, l'avis dudit comité est réputé favorable* ». Ce délai d'un mois peut être **renouvelé une fois** par le président du CESREES, ce qui est fréquemment le cas et les demandes d'informations complémentaires peuvent venir allonger les délais de réponse.

Ainsi, dans un délai d'**un à deux mois** à compter du dépôt initial (hors temps de réponse de l'investigateur), le CESREES rend un **avis écrit** sur le projet. Cet avis porte sur plusieurs éléments clés :

> La **qualité méthodologique** du projet,
> La **nécessité de recourir à des données personnelles**,
> La **pertinence des données demandées** au regard de la finalité du traitement,
> Le **périmètre de données sollicitées**,
> Et, si applicable, la **qualité scientifique globale** du projet.

Une fois l'avis favorable du CESREES obtenu, le dossier est transmis automatiquement à la CNIL par le HDH pour instruction. C'est à cette étape que les aspects réglementaires et juridiques du projet sont examinés, notamment en matière de protection des données personnelles. La CNIL est la seule autorité compétente pour délivrer l'autorisation formelle permettant la mise en œuvre du projet.

NOS CONSEILS / POINTS D'ATTENTION

Après le dépôt sur la plateforme du HDH, le dossier est examiné par le CESREES, qui peut formuler des demandes de précisions (aucune notification automatique n'est envoyée en cas de message). Il est recommandé de désigner plusieurs personnes en charge du suivi des demandes et de mettre en place une vérification quotidienne de la messagerie pour éviter des retards administratifs. Il est important de répondre rapidement à ces sollicitations pour permettre la poursuite de l'instruction jusqu'à l'autorisation de la CNIL.

4.3 L'examen du dossier par la CNIL

Délai estimé de 4 mois (2 mois renouvelable une fois)

La CNIL dispose d'un délai de deux mois, renouvelable une fois (soit quatre mois au total), pour se prononcer sur une demande d'autorisation. Au-delà de ce délai, l'autorisation peut être considérée comme tacitement acceptée. Toutefois, il faut systématiquement en référer au DPO central de l'AP-HP avant de considérer qu'une absence de réponse vaut autorisation tacite. En pratique, l'instruction par la CNIL s'étend généralement sur environ quatre mois. Durant ce temps, la CNIL peut également solliciter des informations complémentaires, ce qui peut allonger le délai réglementaire. Pour pallier cela, il est essentiel que les documents soumis par l'investigateur soient clairs, complets et cohérents. À noter également que la CNIL joue un rôle de conseil et d'accompagnement auprès des investigateurs. Il est toutefois important de fournir les livrables les plus complets (Cf étape 4.2 La constitution du dossier réglementaire) pour limiter les aller-retours avec la CNIL.

A l'issue des quatre mois, n'hésitez pas à relancer la CNIL afin qu'elle se prononce sur votre demande d'autorisation. En l'absence de retour de la CNIL et si vous avez une question / un doute sur votre dossier, consultez nos DPO.

Parallèlement à ces échanges, le HDH informe la CNAM du projet d'appariement et des démarches en cours. Une fois l'autorisation obtenue, celle-ci est transmise à la CNAM, qui déclenche alors la procédure de conventionnement.

NOS CONSEILS / POINTS D'ATTENTION

Veillez à assurer une cohérence parfaite entre les documents soumis à la CNAM et ceux soumis à la CNIL. En particulier, l'autorisation de la CNIL doit être strictement alignée avec l'expression de besoin, car cette dernière constitue la base sur laquelle les juristes de la CNAM établiront la convention.

Etape 5 : Signer la convention

Une **convention** est établie entre la **CNAM** et le **responsable de traitement (AP-HP)** afin de définir les **modalités d'extraction des données** dans le périmètre autorisé. La phase de conventionnement, dont la durée varie généralement entre **4 et 6 mois**, dépend de la **complexité du projet** ainsi que du **traitement des données** demandées. Il s'agit d'un délai indicatif sur la base des entretiens menés dans le cadre de la rédaction de ce guide. Il n'existe pas de délai réglementaire pour la mise en place de la convention de la CNAM.

Pour rappel, le responsable de traitement (AP-HP) est celui qui détermine la finalité du traitement. En d'autres termes, cela signifie qu'il faut **déterminer clairement le statut des Institutions impliquées** et les responsabilités respectives de chacun dans le cadre du projet de recherche et des actions concernant les données[7].

Etape 6 : Suivre la formation SNDS

Après avoir obtenu l'avis favorable du **CESREES** et de la **CNIL**, une notification est adressée à l'investigateur afin qu'il désigne les personnes appelées à accéder et exploiter les données du SNDS (datascientists, biostatisticiens, pharmaco-épidémiologistes, etc.). Ces personnes devront suivre une **formation obligatoire** pour obtenir l'habilitation nécessaire à l'accès aux données du SNDS. Le calendrier des sessions de formation est disponible sur demande[8].

L'inscription suit un **protocole spécifique** et la formation se déroule en **deux étapes** :

- **Module e-learning** : *Architecture et données* (durée : 7 heures)
- **Session en présentielle** : *Données d'extraction pour accès sur projets* (durée : 3 jours)

Le coût de la formation est estimé à environ **1200 €**. L'habilitation obtenue **n'est pas valable indéfiniment** : une actualisation peut être nécessaire selon l'évolution des projets. La CNAM recommande ainsi d'**ajuster le moment de la formation en fonction de la disponibilité réelle des données**.

[7] Voir le site de la CNIL.

[8] snds.cnam@assurance-maladie.fr

NOS CONSEILS / POINTS D'ATTENTION

Il est fortement recommandé de **coordonner le calendrier de formation avec la date prévisionnelle de mise à disposition des données**. En effet, un délai trop long (par exemple 6 mois) entre la formation et l'accès aux données peut entraîner une perte des réflexes acquis lors de la formation. Une bonne pratique consiste à **attendre environ deux mois après l'avis du CESREES** avant d'effectuer la formation, afin de rester opérationnel au moment de l'accès aux données.

Etape 7 : Réaliser techniquement l'appariement

À ce jour, seul un appariement indirect est possible entre les données de l'EDS de l'AP-HP et celles du SNDS. Ce type d'appariement repose principalement sur les données du PMSI, considérées comme les plus fiables pour effectuer un rapprochement entre les deux bases, à l'échelle du patient.

Pour initier la procédure, un sous-ensemble des **données du PMSI[9] est extrait de l'EDS, accompagné d'un identifiant unique par patient, spécifique au projet**. Ces données d'appariement sont filtrées, préparées puis transmises à la CNAM, qui les compare à la version nationale du PMSI. Si nécessaire, des données complémentaires peuvent être demandées.

La plateforme d'appariement de la **CNAM exploite ces données pour ré-identifier les patients** dans le PMSI national. Cela permet ensuite d'extraire, dans le SNDS, l'ensemble des données demandées pour le projet. Ces données sont mises à disposition dans un espace de traitement sécurisé, accolées à l'identifiant unique du projet. Le reste des données médicales de l'EDS sont enfin déposées sur cette plateforme, avec le même identifiant. L'appariement final entre les deux jeux de données médicales se fait donc sur cet identifiant unique du projet, créé sur l'EDS et reconstruit par la CNAM.

[9] Dans le cas des consultations qui n'apparaissent pas dans le PMSI, l'appariement pourra se faire sur le mois et l'année de naissance, le sexe, le Finess, les dates de chaque première consultation par unité médico-administratives et les actes. Cette possibilité doit toutefois être testée.

7.1 Préparation des données nécessaires de l'EDS sur l'espace de travail de l'AP-HP

Cette étape est essentielle pour s'assurer que l'appariement pourra être effectué correctement sur la plateforme de la CNAM. Pour cela, trois actions doivent être réalisées :

- Sélectionner les variables nécessaires pour l'appariement en s'appuyant sur le dictionnaire des variables :
  - Patient : date de naissance, sexe, département de résidence ;
  - Séjour : dates de début et de fin du séjour, numéros FINESS juridiques et géographiques ;
  - Soins : codes CIM-10 des diagnostics principaux, reliés et associés, codes CCAM des actes, GHM/GHS
- Génération d'un identifiant unique (ID_temp) pour chaque patient de la cohorte (variable incrémentale)
- Analyse de la qualité de la table d'appariement par les datascientists de l'URC :
  - Gestion des valeurs manquantes : une analyse précise des valeurs manquantes et aberrantes dans les variables utilisées pour l'appariement permet de mieux estimer le taux d'appariement final,
  - Identification et suppression des doublons,
  - Vérification de la correspondance entre FINESS géographique et FINESS juridique,
  - Contrôle de la cohérence des variables transmises.
- Il convient de transmettre une définition précise des variables transmises pour l'appariement afin d'éviter tout risque de confusion. Par exemple, la colonne FINESS est ambigüe puisqu'elle peut concerner le FINESS géographique ou le FINESS juridique.

De façon générale, cette étape conduit à un compromis entre le taux d'appariement et la quantité de données filtrées. Des critères de filtrage stricts conduisent à un appariement plus rapide, avec un meilleur taux et à des paires plus fiables, mais ces critères conduisent également à réduire la taille de la cohorte.

Une analyse préalable de la qualité de la table d'appariement permet de minimiser les retours chronophages entre l'EDS et la CNAM ou le HDH.

7.2 Transmission des données sur la plateforme d'appariement

Une fois la table d'appariement finalisée dans l'espace de travail Jupyter de l'AP-HP, il convient de contacter le support de l'EDS (id.recherche.support.dsn@aphp.fr) pour procéder à son extraction. La table d'appariement peut être transmise via l'une ou l'autre des plateformes suivantes :

> La plateforme de la CNAM via un canal sécurisé Bluefiles
> La plateforme du HDH via un protocole HTTPS

Si les deux choix sont possibles, ce guide privilégie un envoi direct à la plateforme de la CNAM. En effet, le HDH ne disposant pas des données du SNDS, il va nécessairement transmettre ces données à la CNAM pour réaliser l'appariement. [10]

7.3 Appariement des données SNDS avec les données de la table d'appariement

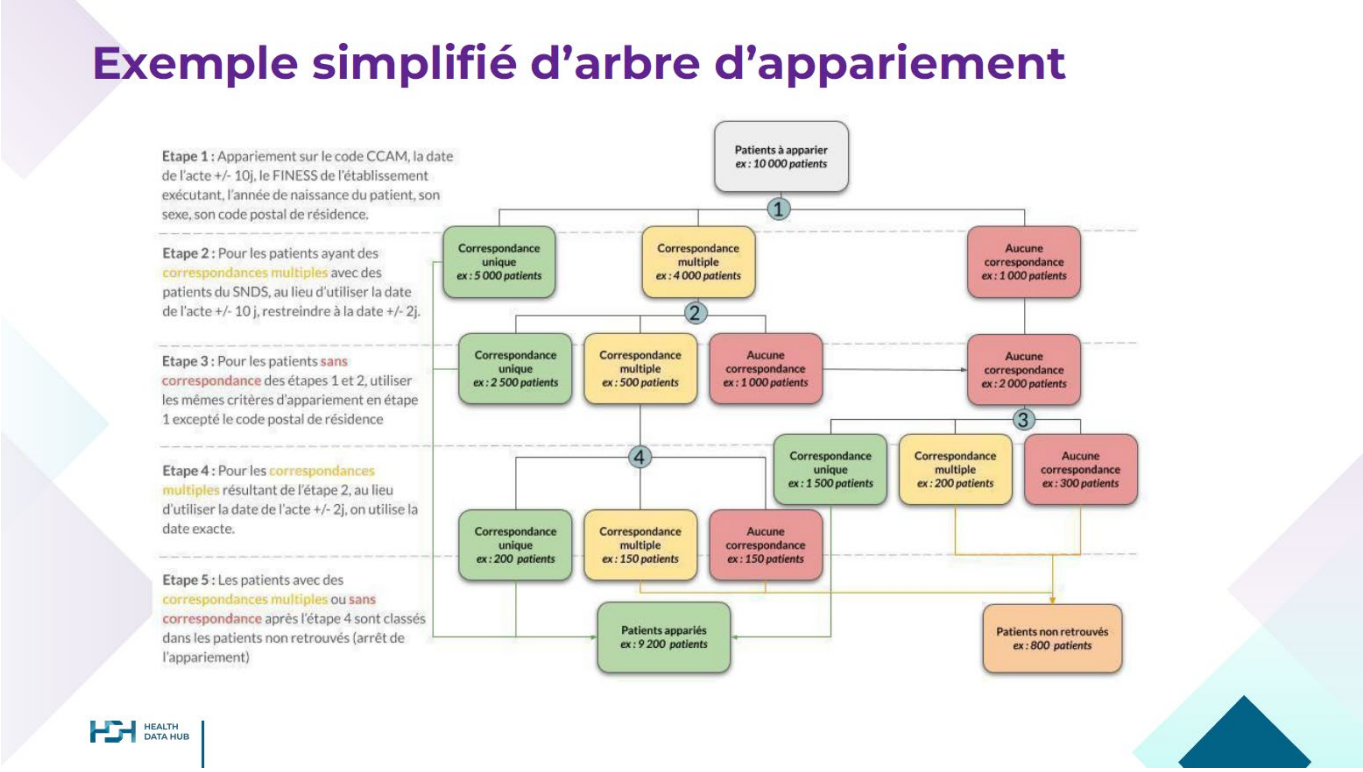
L'appariement indirect s'effectue sur la plateforme de la CNAM. Il peut être réalisé soit par les experts de la CNAM, soit par un datascientist de l'AP-HP ayant suivi la formation SNDS. En règle générale, il est recommandé de confier cette tâche à la CNAM, dont les experts possèdent une expérience plus approfondie. Toutefois, si le datascientist dédié au projet maîtrise bien le SNDS et est habitué à la plateforme, il peut proposer de prendre en charge l'appariement lui-même ce qui limite la charge de la CNAM. Ce choix doit être précisé dès la rédaction de l'expression de besoin.

Techniquement, ce processus se déroule en plusieurs étapes, souvent représentées sous la forme d'un arbre d'appariement :

> Premier appariement
  Il repose sur les variables les plus fiables telles que l'année et le mois de naissance, le sexe, les FINESS juridique et géographique, les dates de début et de fin de séjour, ainsi que les diagnostics principaux et associés. Cette étape, qui se déroule généralement en plusieurs sous-étapes, permet d'identifier environ 85 % à 90 % des patients. Le taux d'appariement varie en fonction de la typologie des séjours PMSI (les séjours ambulatoires offrant moins d'informations), des dates de séjour, et de la concordance entre les données de l'EDS et du SNDS sur les dates de séjour[11].
> Deuxième appariement
  Cette étape intègre les diagnostics associés supplémentaires et les actes médicaux. En fonction de la qualité de la table d'appariement, elle permet de récupérer une proportion variable des patients restants.

[10] Une autre option pourrait être de proposer à la CNAM une plateforme de traitement certifiée HDS qui pourrait accueillir les données (possibilité à vérifier avec la CNAM et à mentionner impérativement dans l'expression des besoins).

[11] Des travaux en cours sur la base centrale de l'EDS visent à prendre davantage en compte la base AREM (transmise à la CNAM) dans les données mises à disposition sur les espaces projets. Cela devrait conduire à une meilleure adéquation entre les variables de l'EDS et celles de la CNAM.



Source : HDH

Les résultats de l'appariement vous sont présentés, avec la possibilité de réaliser plusieurs itérations jusqu'à l'atteinte d'un taux d'appariement satisfaisant. Pendant ce processus, les experts en charge de l'appariement contacteront l'investigateur du projet pour l'informer de l'avancement, pour demander des données et des informations complémentaires sur la définition des variables en cas de besoin. Il est préférable d'indiquer une liste de variables d'appariement assez large dans le dossier CNIL car, si elles n'y apparaissent pas, les variables supplémentaires ne pourront pas être utilisées.

NOS CONSEILS / POINTS D'ATTENTION

> Les appariements réalisés sur des données plus anciennes sont généralement moins fiables, notamment parce que chaque établissement de l'AP-HP possède, depuis 2016, un code FINESS distinct, et non un code unique pour l'ensemble de l'AP-HP.
> Il est essentiel de supprimer soigneusement les doublons avant de transmettre les données à la CNAM.

« Sur un projet, on a réussi à chaîner 500 000 patients avec un taux d'appariement satisfaisant » - Investigateur

7.4 Transfert des données médicales sur la plateforme de traitement

Comme indiqué ci-dessus, la dernière étape du processus d'appariement est la **transmission des jeux de données médicales de l'EDS et du SNDS sur la plateforme d'appariement** avec un identifiant commun reconstruit grâce aux données du SNDS. Des données médicales individuelles peuvent être transmises sur la plateforme d'appariement, mais il ne peut s'agir que de données structurées. En particulier, **les données textuelles et les images restent sur l'espace projet de l'AP-HP**. Si une analyse de données non-structurées (NLP ou analyse d'images) est nécessaire à la réalisation du projet, cette dernière doit être réalisée en amont du transfert vers la plateforme d'appariement.

Ce guide recommande de traiter les données sur la plateforme de la CNAM qui dispose d'un historique de projets plus important et qui constitue donc un choix plus sûr. Toutefois, cette plateforme de traitement est limitée dans les langages de programmation proposés, seul SAS et R étant disponibles. De plus, le nombre de packages est limité sur la plateforme de la CNAM et les demandes d'installation sont soumis à de longs délais. Il convient d'anticiper au maximum cette limitation en réalisant, sur les espaces projet de l'AP-HP, autant de pré-traitement que possible afin de limiter les analyses sur la plateforme de traitement de la CNAM à des statistiques descriptives et des modèles statistiques simples (excluant par exemple l'entraînement de modèles de *machine learning*).

Une alternative possible est l'utilisation de la plateforme de traitement du HDH[12]. Cette plateforme propose d'autres langage de programmation comme python mais dispose d'un historique d'utilisation plus faible. À noter que l'utilisation de cette plateforme nécessite un travail d'homologation de la part des porteurs du projet[13].

L'équipe de l'EDS vous assiste dans le transfert des données médicales de l'EDS vers la plateforme de traitement (HDH ou CNAM). Les données sont ensuite appariées par le **gestionnaire de la plateforme de traitement** et prêtes à être exploitées.

NOS CONSEILS / POINTS D'ATTENTION

Le choix de la plateforme de traitement des données est important. Il doit être guidé par les compétences techniques des datascientists en charge de l'analyse : si ces derniers sont plus familiers avec un langage en particulier, cela influencera le choix. À titre d'exemple, la plateforme du HDH repose sur Python, tandis que celle de la CNAM privilégie R et SAS. Attention toutefois au caractère récent de la plateforme du HDH qui a pu, pour des projets antérieurs, conduire à des difficultés techniques pour la réalisation du traitement.

[12] Pour plus d'information, voir la documentation de cette plateforme.
[13] Pour plus d'informations, voir cette question dans le forum du HDH et la page 6 de la documentation.

NOTRE OFFRE D'ACCOMPAGNEMENT INNOVATION ET DONNEES

VOS DÉFIS

Mener un projet de chaînage implique plusieurs contraintes

- Des délais longs et peu maîtrisables : plusieurs mois peuvent s'écouler entre le montage du dossier et l'analyse des données appariées
- Une procédure réglementaire exigeante, mobilisant CESREES, CNIL, HDH et CNAM, avec un besoin de connaissance fine
- Une forte technicité méthodologique notamment pour construire une table d'appariement fiable et exploitable
- Un risque important de va-et-vient et de délais prolongés si les données transmises sont incomplètes, incohérentes ou mal préparées

NOS SOLUTIONS

L'équipe EDS propose un accompagnement dédié pour sécuriser chaque étape du processus et maximiser vos chances de succès

> Un interlocuteur privilégié pour vous accompagner tout au long de votre projet

> Un appui méthodologique et technique pour définir les variables nécessaires au chaînage et créer la table d'appariement (extraction, pseudonymisation, création d'identifiants uniques)

> Des échanges facilités avec la CNAM et/ou le HDH pour la transmission de la table via les canaux sécurisés

Le coût de la prestation est précisé lors de la réunion CSSI, accompagnée d'une étude de faisabilité détaillant le budget. Il s'élève généralement à environ 3 500 € le temps consacré au transfert des données vers les plateformes du HDH ou de la CNAM.

ANNEXE

Projets passés de chaînage EDS / SNDS

OPPORTOS

Prediction des fractures par depistage opportuniste de l'osteoporose

Investigateur : Dr Christian Roux

EXONEPHRO

Intérêt clinique de la réalisation d'un examen génétique de type Whole Exome Sequencing devant une néphropathie de cause inconnue ou incertaine

Investigateur : Dr Olivier Steichen

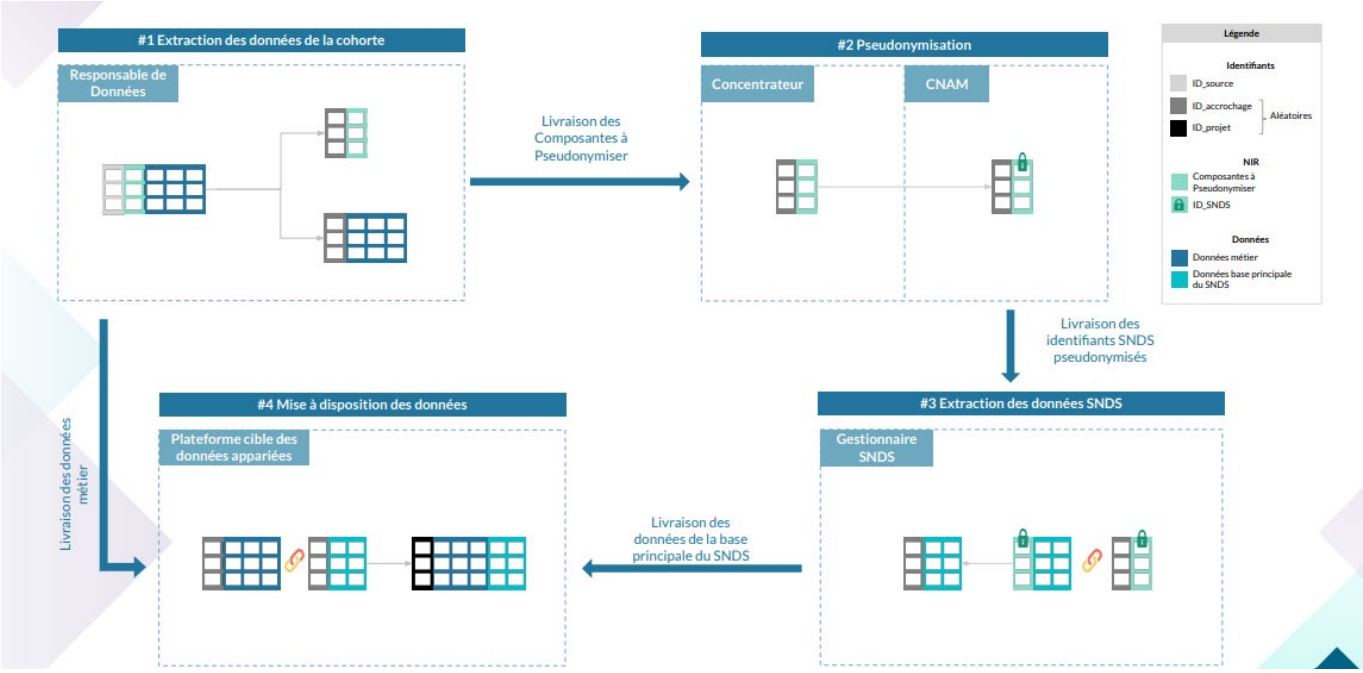
EPI2

Etude épidémiologique de l'épilepsie en France avec validation d'un algorithme de ciblage de l'épilepsie dans le SNDS

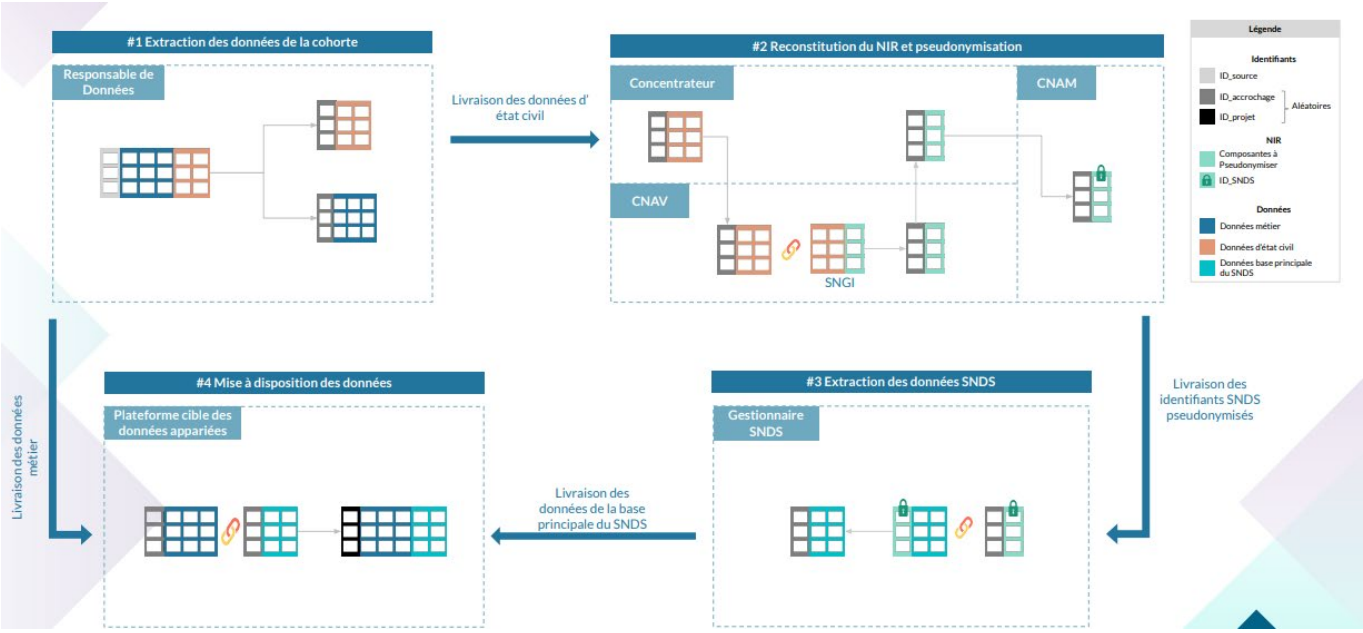
Investigateur : Dr Vincent Navarro

Les trois circuits d'appariement possible

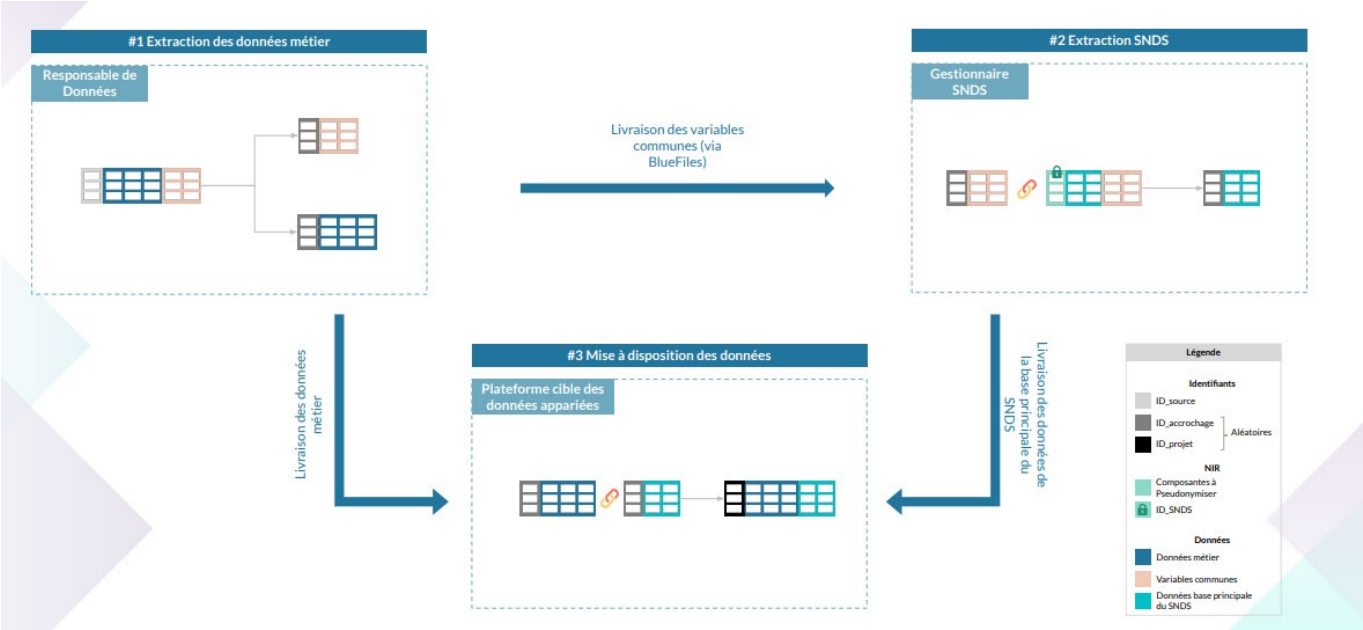
Appariement direct via le NIR



Appariement avec reconstitution du NIR



Appariement indirect



Liste des actions à réaliser pour le chaînage des données avec le SNDS

Actions	Responsable	
Prendre contact avec l'URC		
Contacter le coordinateur EDS de l'URC	Investigateur du projet	☐
Préparer le kit de demande d'accès aux données EDS	URC / Investigateur du projet	☐
Préparer le dossier juridique : demande d'avis au CESREES, autorisation CNIL	URC / Investigateur du projet	☐
Enregistrer la demande dans Timetonic (formulaire en ligne)	URC	☐
Réaliser une première étude méthodologique et scientifique du projet	URC	☐
Etudier la faisabilité technique du projet		
Solliciter l'équipe EDS pour une étude technique complémentaire (CSSI)	Investigateur du projet / URC	☐
Organiser réunion CSSI pour discuter des contraintes (ressources, calendrier, réglementaires, accès aux données, appariement)	Equipe EDS (RUPP)	☐
Valider l'étude de faisabilité avec chiffrage des coûts	Investigateur du projet	☐
Utiliser cette étude pour demander un financement adapté si besoin	Investigateur du projet	☐
Accéder aux données de l'EDS		
Présenter le projet au CSE de l'AP-HP avec intention de chaînage SNDS	Investigateur du projet / URC	☐
Attendre l'avis favorable du CSE	/	☐
En cas de partenariat externe, engager des discussions contractuelles via la DRCI	Investigateur du projet / URC / partenaire	☐
Construire la requête de sélection patients, préparer le	Investigateur du projet / URC	☐

document de traçabilité et la matrice d'habilitation et l'envoyer à l'équipe EDS		
Livraison des données pseudonymisées dans l'espace sécurisé Jupyter	Equipe EDS (ISD)	☐
Validation réglementaire par CESREES et CNIL		
Constituer le starter kit du HDH	Investigateur du projet / URC	☐
Déposer le dossier complet sur la plateforme du HDH	Investigateur du projet / URC	☐
Attendre l'examen du dossier par le CESREES et répondre aux demandes de compléments	/	☐
Après avis favorable CESREES, dossier transmis à la CNIL pour instruction	CESREES	☐
Attendre l'avis CNIL	/	☐
Après avis favorable de la CNAM, transmission de l'avis à la CNAM	CNIL	☐
Signature de la convention avec la CNAM		
Signer la convention avec la CNAM définissant modalités d'extraction des données dans périmètre autorisé	Investigateur du projet / URC / DRCI	☐
Suivre la formation SNDS		
S'inscrire à la formation	Investigateur du projet / URC	☐
Réaliser module e-learning « Architecture et données »	Investigateur du projet / URC	☐
Participer à la session « Données d'extraction pour accès sur projets »	Investigateur du projet / URC	☐
Réaliser techniquement l'appariement		
Préparer la table d'appariement dans l'espace projet Jupyter (sélection des variables clés, génération d'un identifiant unique patient	Investigateur du projet / URC	☐

Analyser la qualité de la table d'appariement (gestion valeurs manquantes, suppression doublons, cohérence FINESS, contrôle variables)	Datascientist du projet	☐
Valider la table d'appariement	Equipe EDS (ISD)	☐
Contacter l'équipe support de l'EDS (RUPP) pour transférer la table d'appariement	Datascientist du projet	☐
Transmettre la table d'appariement à la plateforme CNAM ou HDH	Equipe EDS (RUPP)	☐
Apparier les données SNDS avec les données de la table d'appariement	CNAM	☐
Transférer les données médicales de l'EDS vers la plateforme traitement	Equipe EDS (RUPP)	☐