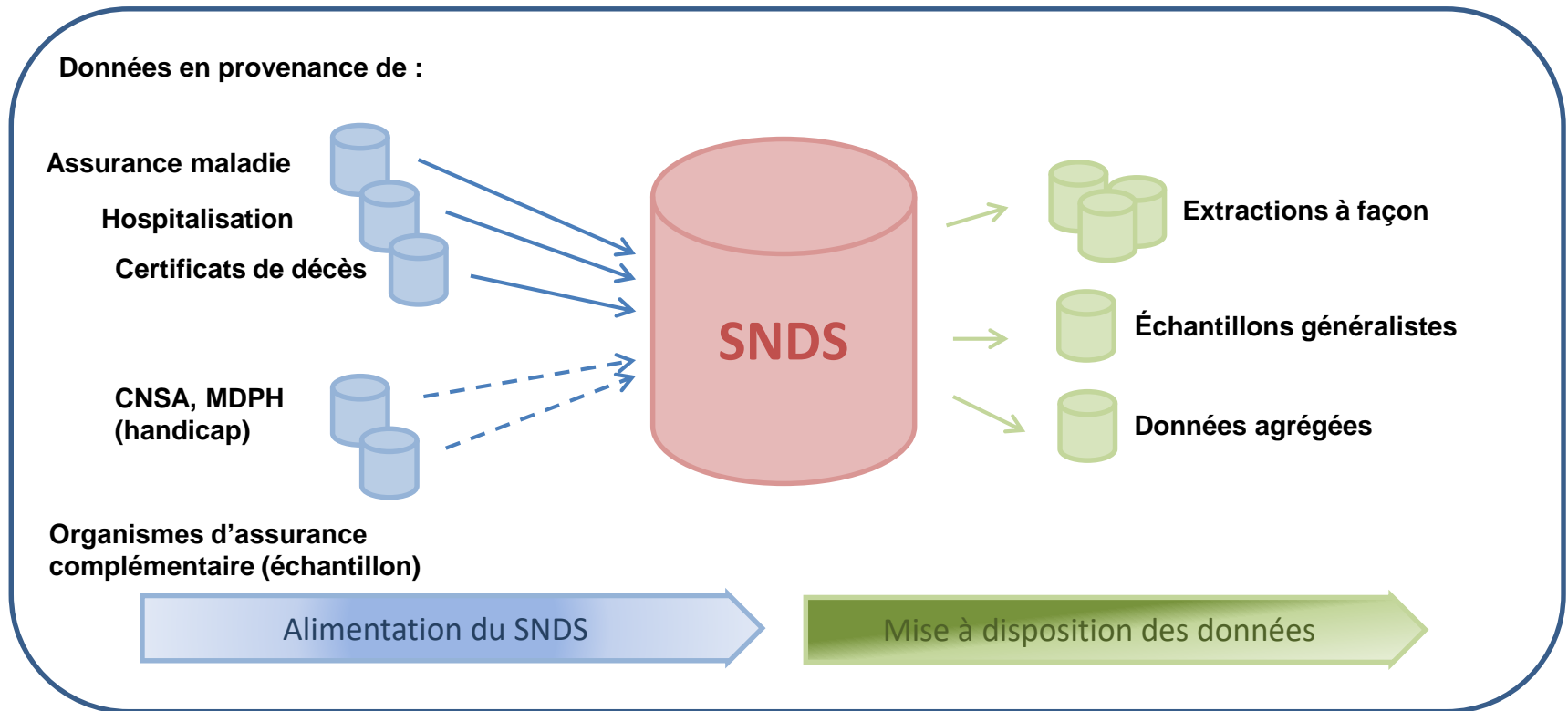


# LE SYSTÈME NATIONAL DES DONNÉES DE SANTÉ (SNDS)

## Qu'est ce que le SNDS ?



Les données de tous les citoyens y sont déversées : données pseudonymisées, conservées 20 ans

1,2 Md de feuilles de soins et 11 millions de séjours hospitaliers alimentent le SNDS annuellement → 450 To gérés par la Cnam

## Quelles données ?

- **Sur les patients :**  
Mois et année de naissance, sexe, notion de CMU-C, diagnostic de l'ALD (affection de longue durée), commune et département de résidence, date de décès et la cause médicale de décès
- **Sur la consommation de soins en ville :**  
Toutes les prestations remboursées par date de soins avec le codage détaillé de la prestation (actes médicaux avec le code CCAM, biologie, dispositifs médicaux, code CIP des médicaments),
- **Sur la consommation de soins en établissement :**  
Les données du PMSI (MCO-SSR-HAD-PSY) pour l'ensemble des établissements sanitaires (motif médical, actes CCAM, durée de séjour, mode de sortie,...)  
Les médicaments et dispositifs facturés « en sus » et l'activité externe des hôpitaux (depuis 2009)

## Quelles données ?

### Sur les pathologies traitées

- à partir des diagnostics : mise en ALD, hospitalisations
- à partir des consommations : médicaments ou actes ou dispositifs médicaux traceurs

→ besoin d'algorithmes pour identifier les pathologies (incidence, prévalence ...)

### Des limites structurelles

- pas de motif de recours aux soins, de résultat d'examens, de notion d'antécédents personnels ou familiaux,
- pas de données sociales (CMU uniquement) ou environnementales,
- pas d'informations sur des possibles facteurs de risque de survenue de certaines pathologies (ex risque cardiovasculaire augmenté chez les fumeurs)

→ possibilité de dépasser ces limites : traitements statistiques, appariements, ...

## Qui peut accéder au SNDS ?

Les acteurs **publics et privés** selon deux modalités différentes :

- L'accès permanent à certaines catégories de données accordé aux organismes disposant d'une mission de service public
  - L'accès sur projet pour tous les acteurs publics et privés (autorisation de la CNIL)
- + un principe d'open data pour tous les jeux de données ne présentant aucun risque de réidentification (anonymes).

# L'EDP SANTÉ

-

**Une source pour les études croisant données  
sociales et données de santé**

# L'EDP Santé : un projet de rapprochement de sources administratives

## Échantillon Démographique Permanent

Base longitudinale enrichie chaque année par l'Insee.

Il permet de réaliser des études individuelles sur les **trajectoires démographiques, géographiques,**

**professionnelles, ou économiques.**

Accès interne pour la DREES, au CASD pour les chercheurs



## Système National des Données de Santé

Base longitudinale gérée par la CNAM.

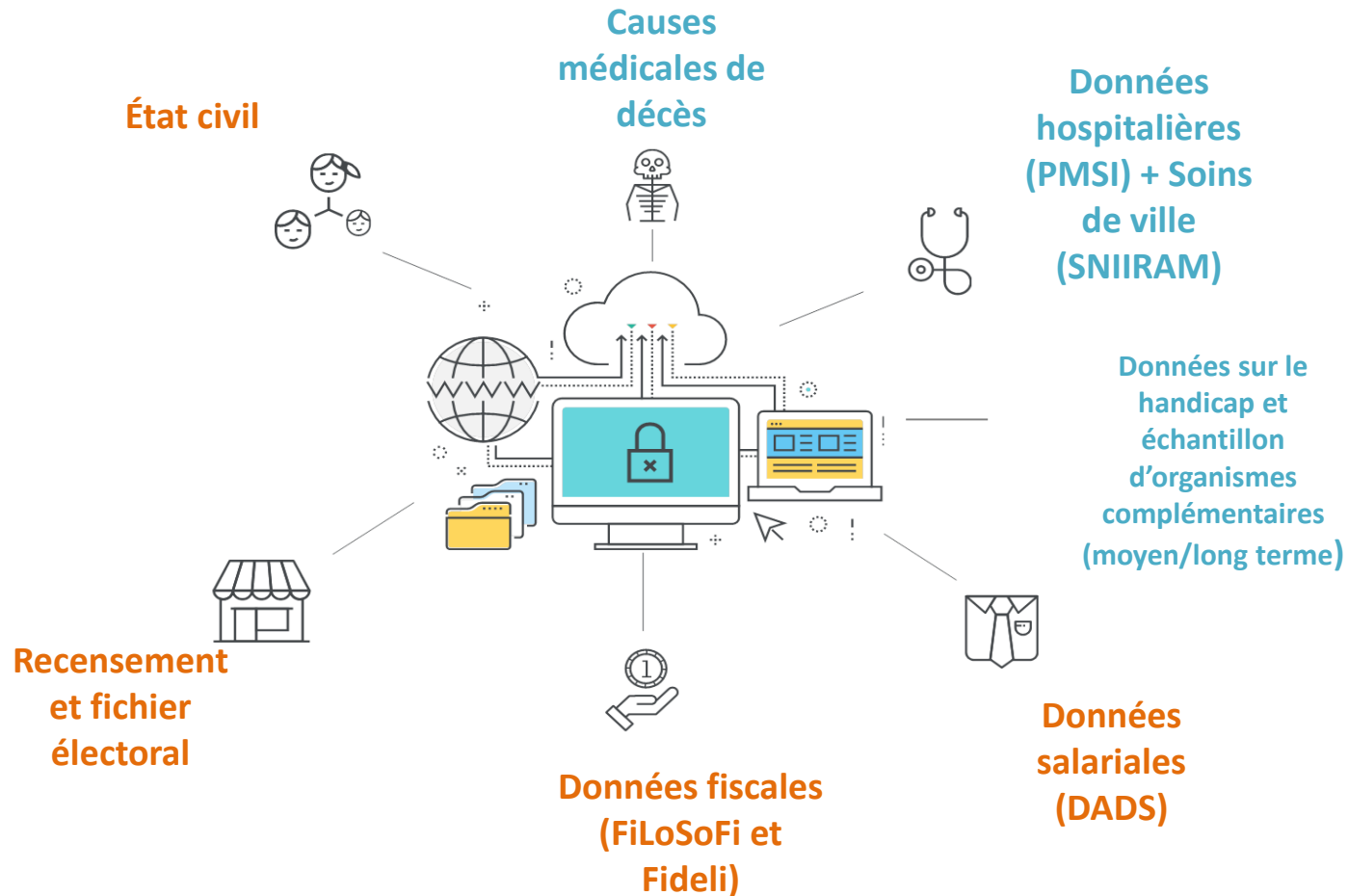
Il permet de réaliser des études sur les **parcours individuels de soin et de santé.**

Accès via un portail pour la DREES et les chercheurs

### Objectif de l'appariement

Obtenir une source inédite permettant de croiser sur un échantillon représentatif des données socio-économiques et des données médico-administratives sur la santé et le recours aux soins.

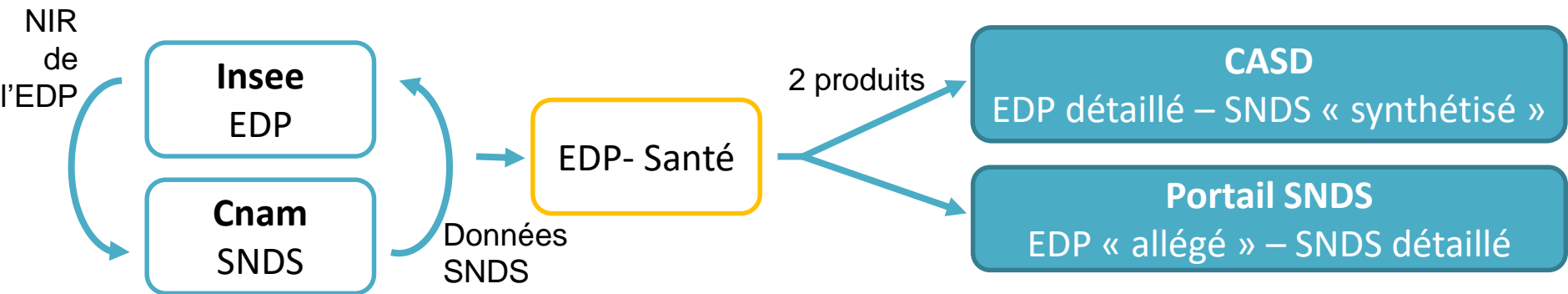
L'EDP Santé c'est **3,6 millions d'individus** et **9 sources** rapprochées.





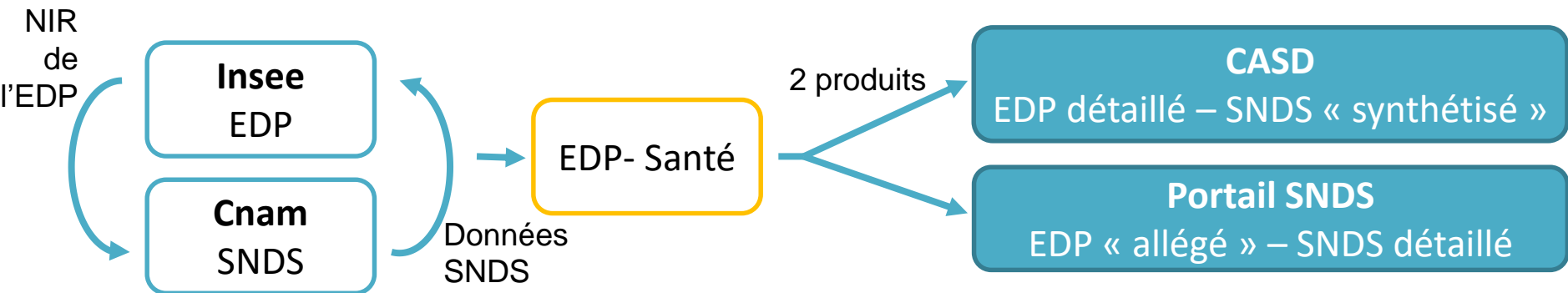
## Des enjeux **juridiques** et **techniques** forts dans la réalisation du projet

### Circuit de données prévisionnel



Des enjeux **juridiques** et **techniques** forts dans la réalisation du projet

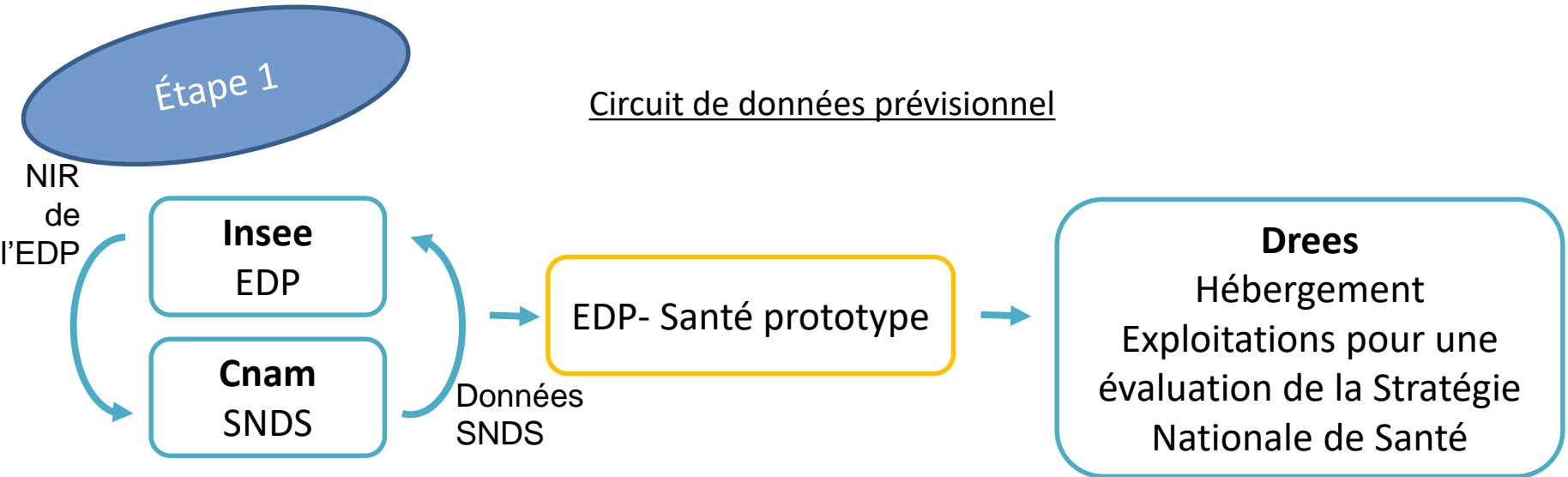
Circuit de données prévisionnel



Obstacle juridique : la loi ne permet pas la constitution d'entrepôts pérennes contenant des données du SNDS et d'autres données appariées

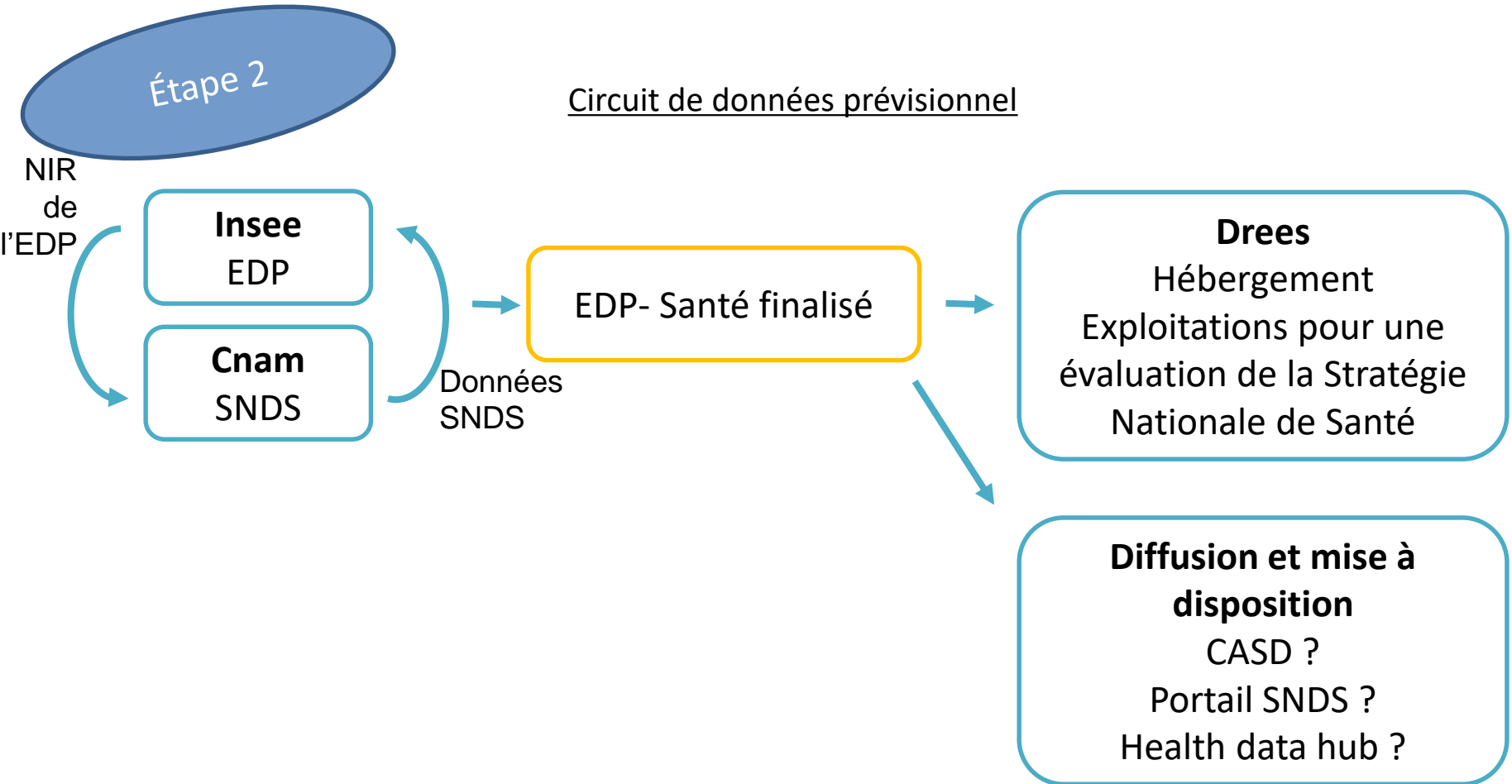
Solution : échelonner le projet en deux étapes

Des enjeux **juridiques** et **techniques** forts dans la réalisation du projet



**Objectif de réalisation de la 1<sup>e</sup> étape du  
projet : 1<sup>er</sup> trimestre 2019**

## Des enjeux **juridiques** et **techniques** forts dans la réalisation du projet



## Un éventail **d'études** et **problématiques** très large

- Déterminants sociaux et territoriaux de la santé et du recours aux soins  
Situation familiale, situation socio-économique, situation professionnelle et travail, trajectoires spatiales et leurs liens avec la santé
- Système de protection sociale et d'assurance maladie  
Poids des dépenses de santé selon la situation sociale, et répartition du reste à charge après assurance maladie obligatoire, et complémentaire (moyen terme)
- Handicap et dépendance (moyen terme)  
Liens entre la situation sociale et l'entrée en institution, ou la prise en charge du handicap

# LE HEALTH DATA HUB



# HEALTH DATA HUB

MISSION DE PRÉFIGURATION

## Une priorité stratégique pour la France



- **Remise du rapport Villani** le 29 mars 2018 lors de l'événement « AI for Humanity », qui présente la vision de l'IA pour la France et qui identifie **quatre secteurs stratégiques, dont la santé**

---

- Atouts majeurs de la France qui dispose d'un **patrimoine de données de santé d'une grande richesse**, avec le SNDS notamment, et qui est reconnue pour son **excellence en matière de recherche médicale et mathématique**

---

- Enjeux stratégiques
  - Renforcer l'**excellence du système de soins** et développer une médecine de précision, personnalisée pour les patients
  - Valoriser le **patrimoine de données et le tissu industriel français** dans un environnement dominé par les GAFAM
  - Assurer la **souveraineté nationale** pour garantir un usage éthique des données sensibles et la fiabilité des innovations



# La mission de préfiguration

## Trois pilotes



### Marc Cuggia

PUPH Informatique Médicale Biostatistique  
Co-pilote de la mission



### Dominique Polton

Présidente de l'Institut national des données de santé  
Co-pilote de la mission



### Gilles Wainrib

Co-fondateur d'OWKIN  
Co-pilote de la mission

## L'ensemble de l'écosystème mobilisé



### 1 groupe de travail plénier

Instituts de recherche, assurance maladie, ministères, établissements et professionnels de santé, industriels



### 114 auditions

Etablissements et professionnels de santé, chercheurs, associations de patients, acteurs institutionnels, start-ups, représentants de l'industrie des produits de santé et du numérique, assurances, collèges et sociétés savantes



### 5 ateliers thématiques, un benchmark international et une expertise juridique

Attente des industries de la santé, échanges avec les éditeurs de logiciels, conditions du partage de la donnée, architecture technique, échanges avec des représentants d'agences sanitaires

# Un grand potentiel de transformation et d'innovation

## Près de 100 cas d'usage identifiés

### Recherche médicale



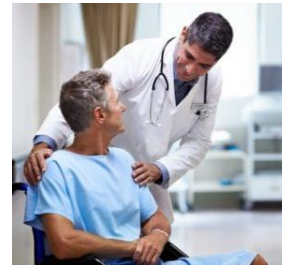
Mettre en évidence des biomarqueurs et des prédispositions génétiques à certains cancers, identification des états précancéreux, accélérer la **recherche clinique**

### Suivi et information des patients



Développer pour les patients des outils d'aide à l'interprétation, par exemple des examens biologiques ou d'**imagerie médicale**

### Appui aux professionnels de santé



Prédire les risques de complications liés à certaines prises en charge (par ex. la chirurgie) et adapter celle-ci aux caractéristiques du patient, (par ex. pour la prise en charge de l'obésité)

### Pilotage du système de santé



Identifier les ré-hospitalisations potentiellement évitables pour améliorer l'offre de soin ; mieux identifier les effets **indésirables** des médicaments

## Les attentes de l'écosystème

---

1

**Mieux connaître** et pouvoir plus **facilement accéder** au patrimoine de données de santé, **décloisonner** les données et **faciliter leur rapprochement**, dans le respect du droit du citoyen

2

Avoir accès à des **moyens adaptés et suffisants** (technologiques, humains et financiers) pour consolider et valoriser le patrimoine de données

3

Fédérer l'écosystème autour d'un **modèle économique d'ensemble** favorisant le partage et reconnaissant les efforts de chacun

# Ambitions



Mettre au service du plus grand nombre les données de santé financées par la solidarité nationale dans le respect de l'éthique et des droits fondamentaux de nos concitoyens

## Pilier #1

Partage de la donnée selon une gouvernance unifiée et garantie par la loi



Evolution du cadre législatif et conduite du changement

## Pilier #2

Allocation de moyens pour consolider le patrimoine de données



Moyens financiers et politique de structuration des données

## Pilier #3

Création d'un tiers de confiance national pour accompagner l'accès et offrir des moyens mutualisés pour valoriser des données dans un environnement sécurisé



Health Data Hub, en lien avec les acteurs de l'écosystème

## Pilier #4

Identification des projets pilotes prometteurs en termes de santé publique et de promotion de la filière industrielle



Moyens financiers (FII, grands défis ?)

# Le Health Data Hub

## Vision d'ensemble

### Rôle du Health Data Hub



#### Moteur et fédérateur

Tiers de confiance fixant une ambition et des objectifs communs, dans le respect du droit des patients et en assurant la transparence avec la société civile



#### Vitrine nationale et internationale

Patrimoine de données et offre de service lisibles pour les acteurs publics et privés de la santé



#### Guichet unique

Guichet unique pour collecter les données et pour faciliter l'accès aux utilisateurs : règles d'accès transparentes, accompagnement dans les procédures de demande d'accès et dans la valorisation économique, documentation des bases partagées, échantillons de test, capacités techniques et juridiques pour les appariements, etc.



#### Garant de la qualité

Diffusion de standards de qualité internationaux pour les données partagées via le Hub et valorisation de l'effort de collecte et de mise en qualité des producteurs



#### Mutualisation de technologies et d'expertises

Plateforme technologique sécurisée, expertises en propre, animation de la communauté et valorisation des compétences et initiatives existantes.

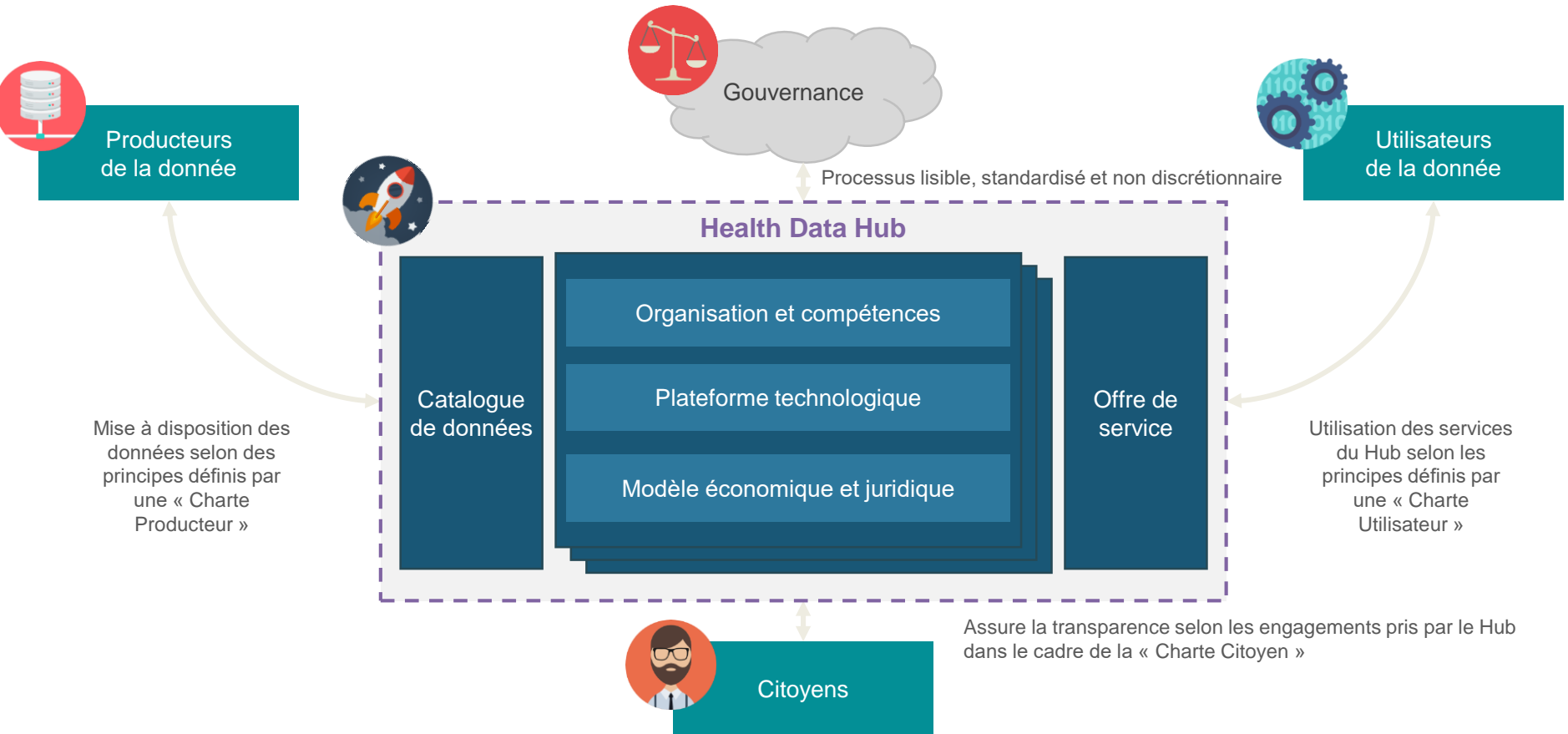


#### Promotion de l'innovation

Développement d'un environnement dans lequel les innovations peuvent prospérer : échantillons de données test, développements d'API, modèle économique favorisant l'innovation (gestion des questions de propriété intellectuelle et de partage du risque...), délais compatibles avec les contraintes des petits et gros industriels

# Le Health Data Hub

## Organisation générale



# Le Health Data Hub

## Gouvernance de la donnée

1

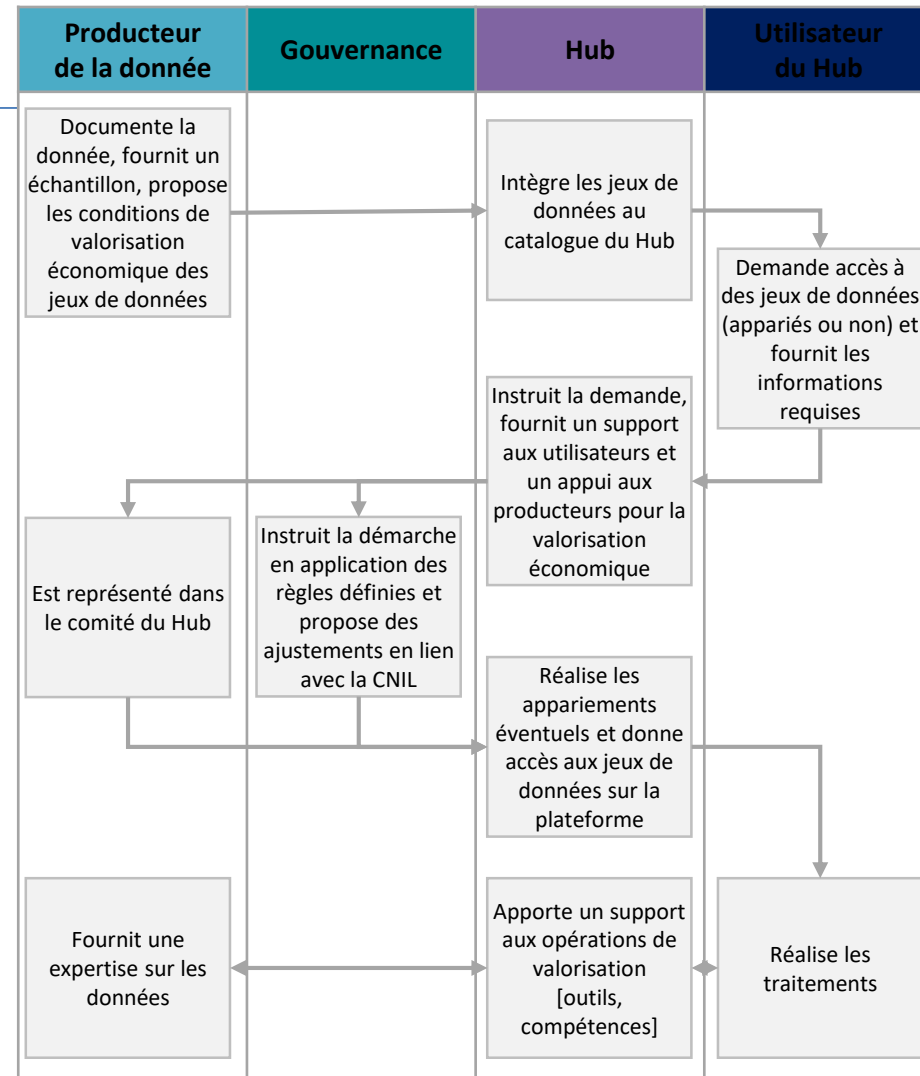
Un **guichet unique** à un catalogue de données en constante évolution. Compte-tenu de la sensibilité des données, un **comité éthique et scientifique** instruit les demandes.

2

Les règles d'accès sont **transparentes et non discrétionnaires**. Les délais sont fixés et les refus explicites. Des représentants des producteurs interviennent pour juger de la faisabilité des projets si nécessaire.

3

En cible, **ce comité et ces règles sont uniques**, notamment apporter de la visibilité, pour faciliter les rapprochements de données et raccourcir les délais d'accès. En pratique, les producteurs de données partenaires du Hub font converger les gouvernances et un comité est mis en place pour les projets appariant plusieurs sources.





# Le Health Data Hub

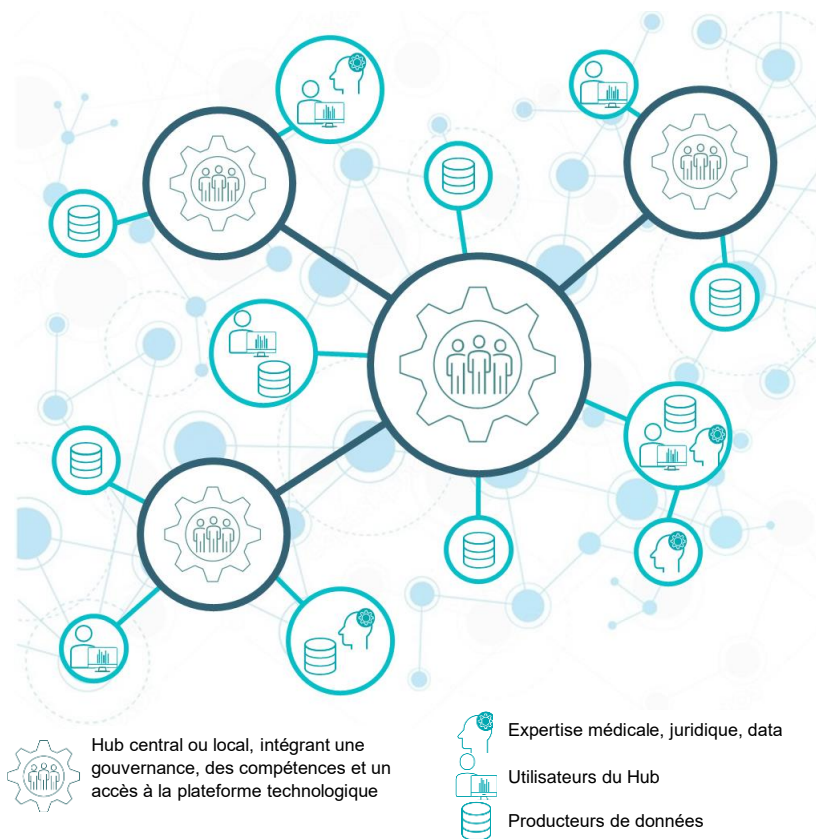
## Offre de service





# Le Health Data Hub

## Organisation en réseau



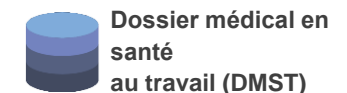
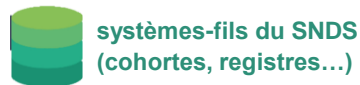
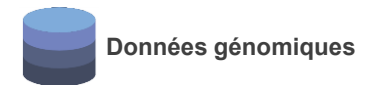
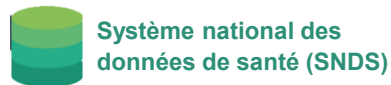
### Principes clé de l'organisation du Health Data Hub

- Un réseau constitué d'un Hub central et de Hubs locaux
- Une gouvernance, une offre de service, des processus et un engagements de résultats communs
- Une mutualisation des investissements technologiques à l'état de l'art et à haut niveau de sécurité

# Le Health Data Hub

## Patrimoine de données

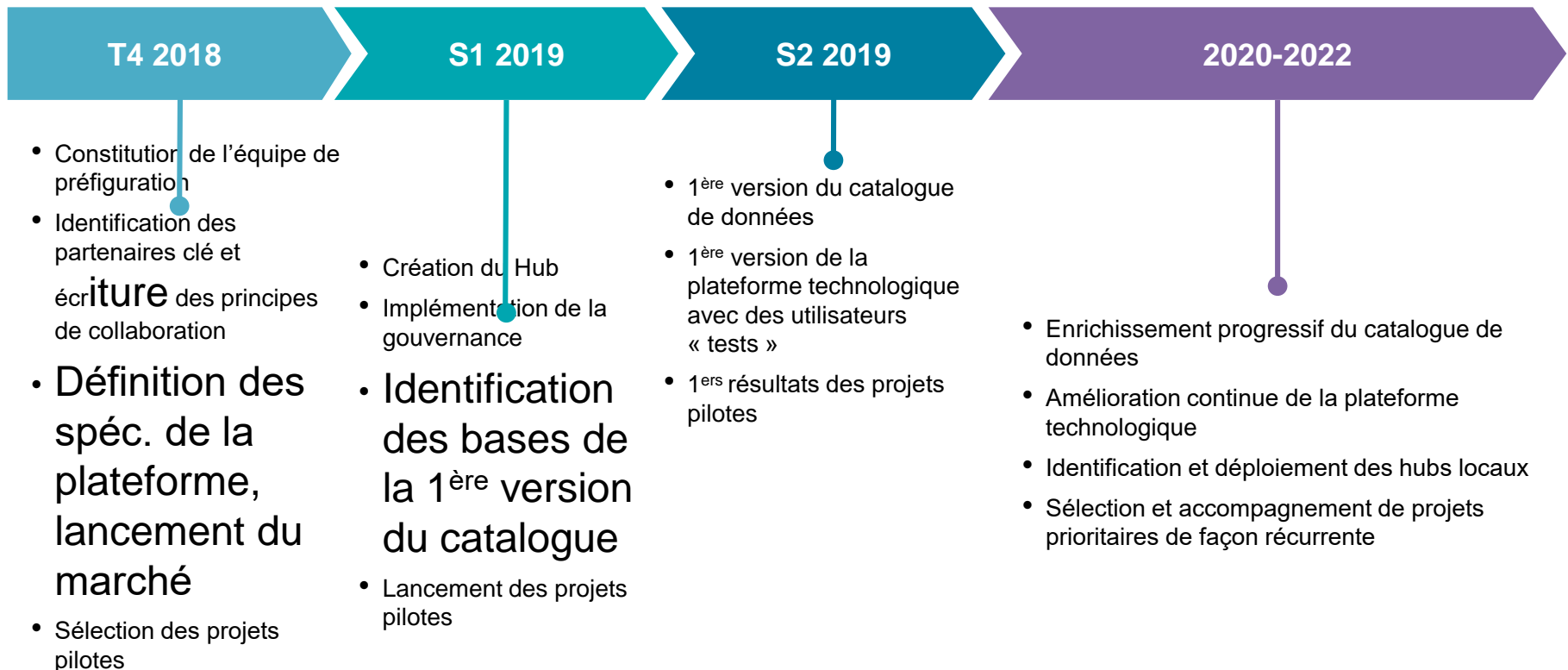
### Constitution progressive du catalogue de données



# Le Health Data Hub

## Feuille de route

### Développement collaboratif, itératif et agile



# Un exemple d'utilisation

# Semeia

- 50% des patients avec une affection longue durée ne suivent plus régulièrement leur traitement après un an (OMS)
- Des études montrent l'efficacité du support personnalisé aux patients



Concentrer les efforts sur les patients ayant les besoins les plus prégnants



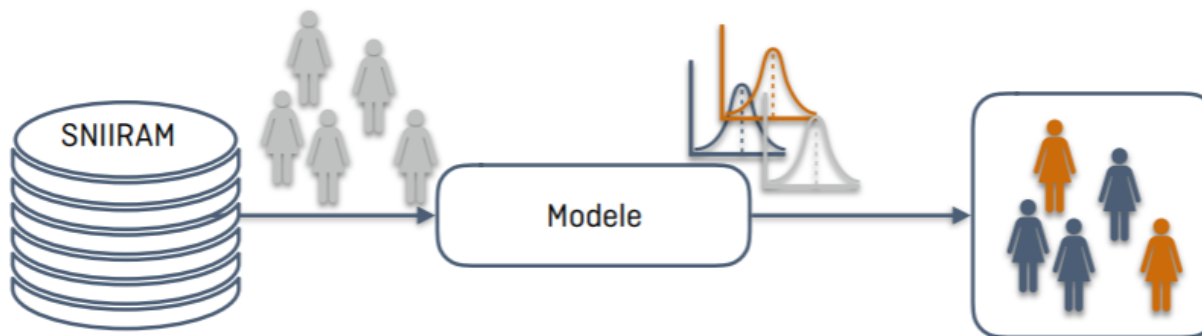
Intervenir **uniquement** aux moments opportuns



Adapter les messages délivrés et le canal de communication au contexte du patient

# Plan : modèles prédictifs appliqués au SNIIRAM

- Les données étudiées : **le SNIIRAM**
- Analyse de phases :
  - Étude préliminaires : **analyse de survie itérative**
  - Comparaison d'approches **d'apprentissage supervisée**
- Analyse dynamique :
  - **Réseaux de neurones récurrents appliqués aux transactions patients**



# Les données étudiées

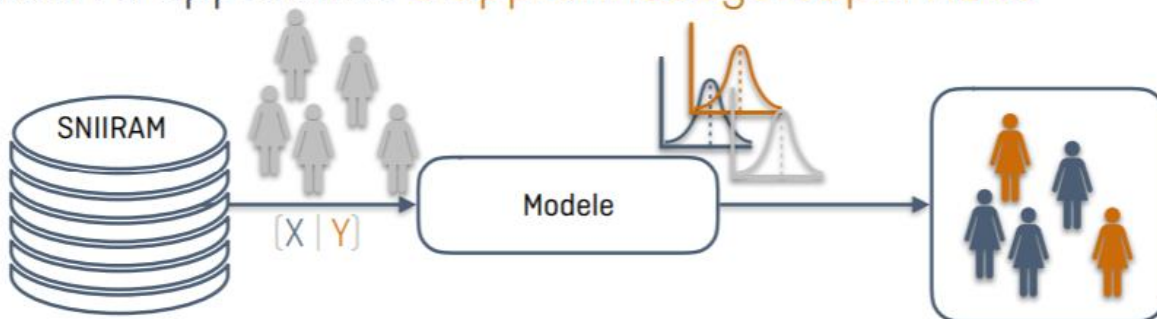
- le **SNIIRAM** : enregistrement des données de remboursements du système de santé français
- Focus sur le **cancer du sein de la femme** : cohorte composée de 50% des femmes, diagnostiquées avec un cancer du sein, ayant acheté du Tamoxifène, entre 2013 et 2015
- Les données sont travaillées pour reconstruire le parcours de soin des patientes, aussi bien en **phases de traitement** qu'en **séquences d'évènements temporels**





# Etude de phases de traitement :

Comparaison d'approches d'apprentissage supervisée



## Différents modèles :

- Régression logistique
- Arbre de décision
- Gradient boosting
- Multi-layer perceptron

