

# La gouvernance des données massives critiques

## Gestion de données génomiques

Juin 2026

PJ Bousquet

---



# Infrastructures du Plan France médecine génomique 2025



FRANCE MEDECINE  
GENOMIQUE 2025



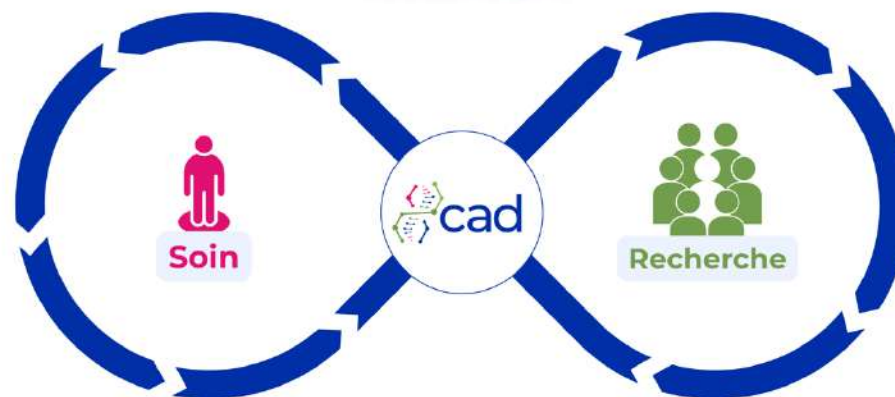
Le Plan France Médecine Génomique (PFMG2025) a pour objectif d'implémenter la Médecine Génomique en France dans un continuum soin-recherche.

Il s'appuie sur 3 piliers principaux :

Un Centre de Référence, d'Innovation, d'expertise  
et de transfert (CRefIX)



Un réseau de laboratoires de  
séquençage génomique (LBM-FMG  
pour *Laboratoire de Biologie Médicale  
France Médecine Génomique*)



Un centre national de  
Collecte et d'Analyse des  
Données (CAD)



« Mise en place d'un **centre national de calcul intensif** (le « CAD », Collecteur analyseur de données », capable de traiter et d'exploiter le **volume considérable de données générées** et d'offrir les premiers services à destination des professionnels de santé **dans le cadre du parcours de soins** (analyses in silico et outils d'aide à la décision en matières diagnostic, pronostic et d'élaboration de stratégies thérapeutiques) »

E Philippe - 2017



→ Création d'un Groupement d'intérêt public en novembre 2022

*Ce travail a bénéficié d'une aide de l'Etat gérée par l'Agence Nationale de la Recherche au titre du Programme d'Investissements d'Avenir portant la référence ANR-21-ESRE-0001*

# Préambule Etude ou Entrepôt de données de santé ?



# Une étude



Une **étude** est :

- Recherche impliquant la personne humaine Information des personnes (RIPH)
- Recherche n'impliquant la personne humaine, une évaluation dans le domaine de la santé (RNIPH)

La finalité du traitement de données est **précise et ponctuelle**

# Un entrepôt de données



Bases de données destinées à être utilisées notamment à des fins de recherches, d'études ou d'évaluations dans le domaine de la santé

- Collecter et disposer de données massives
- Diversité des données : prise en charge médicale, sociodémographiques, données issues de précédentes recherches ...
- Durée de conservation généralement longue (> 10 ans)
- Possibilité d'une alimentation multi-sources

Les finalités de traitement peuvent être multiples et étalées dans le temps



**Avantage** : Conserver les données et les réutiliser de multiples fois

**Inconvénient** : Lourdeur dans la mise en place et le suivi



Seule la constitution de l'entrepôt est couverte par l'engagement de conformité au référentiel « entrepôt de données de santé » de la Cnil

La réutilisation des données de l'entrepôt dans le cadre de recherches, d'études ou d'évaluations dans le domaine de la santé doit faire l'objet des formalités à part

# Données de santé

---



# Une donnée de santé ?



**CNIL.**

Les données à caractère personnel concernant la santé sont les données relatives à la santé physique ou mentale, passée, présente ou future, d'une personne physique (y compris la prestation de services de soins de santé) qui révèlent des informations sur l'état de santé de cette personne.

---

## Données de santé - Cnil



Cette définition comprend donc par exemple :

- les informations relatives à une personne physique **collectées lors de son inscription** en vue de bénéficier de services de soins de santé ou lors de la prestation de ces services : un numéro, un symbole ou un élément spécifique attribué à une personne physique pour l'identifier de manière unique à des fins de santé ;
- les informations **obtenues lors du test ou de l'examen d'une partie du corps ou d'une substance corporelle**, y compris à partir des **données génétiques** et d'échantillons biologiques ;
- les informations concernant une maladie, un handicap, un risque de maladie, les antécédents médicaux, un traitement clinique ou l'état physiologique ou biomédical de la personne concernée (indépendamment de sa source, qu'elle provienne par exemple d'un médecin ou d'un autre professionnel de santé, d'un hôpital, d'un dispositif médical ou d'un test de diagnostic in vitro).

Cette définition permet d'englober certaines données de mesure à partir desquelles il est possible de **déduire une information sur l'état de santé** de la personne.

---

## Des spécificités relatives à la génétique

---



# Un prélèvement réalisé pour une analyse génétique

Le prélèvement est réalisé **expressément** pour une analyse génétique

Article 16-10 du Code civil (modifié par la Loi 2021-2017 du 2 août 2021)

I.- L'examen des caractéristiques génétiques constitutionnelles d'une personne ne peut être entrepris qu'à des fins médicales ou de recherche scientifique. **Il est subordonné au consentement exprès de la personne**, recueilli par écrit préalablement à la réalisation de l'examen



- Il est donc possible d'exploiter les données de génétique dont celles issues du Plan France médecine génomique 2025 à **l'unique condition** que les personnes aient **signé un consentement** autorisant leur réutilisation dans le cadre d'une étude, recherche ou évaluation
- En absence de consentement, il n'est pas possible de réutiliser les données de génomique
- Exclut le principe **initial** de non-opposition

# Nécessité d'apporter une information claire

Article 16-10 du Code civil (modifié par la Loi 2021-2017 du 2 août 2021)

II.-Le consentement prévu au I est recueilli après que la personne a été dûment informée :

1° De la **nature de l'examen** ;

2° De **l'indication de l'examen**, s'il s'agit de finalités médicales, ou de son objectif, s'il s'agit de recherches scientifiques ;

3° Le cas échéant, **de la possibilité que l'examen révèle incidemment des caractéristiques génétiques** sans relation avec son indication initiale ou avec son objectif initial mais dont la connaissance permettrait à la personne ou aux membres de sa famille de bénéficier de mesures de prévention, y compris de conseil en génétique, ou de soins ;

4° De **la possibilité de refuser la révélation des résultats** de l'examen de caractéristiques génétiques sans relation avec l'indication initiale ou l'objectif initial de l'examen ainsi que des risques qu'un refus ferait courir aux membres de sa famille potentiellement concernés, dans le cas où une anomalie génétique pouvant être responsable d'une affection grave justifiant de mesures de prévention, y compris de conseil génétique, ou de soins serait diagnostiquée.

# En pratique – Données incidentes

Le responsable de l'EDS doit être en **capacité de revenir vers la personne** (de façon directe ou indirecte selon les modalités de constitution de l'EDS)

Article 5.5 référentiel EDS :

Les données directement identifiantes mentionnées au point 5.2.1.1 ne peuvent être réunies dans l'entrepôt que pour les finalités suivantes : (...)

- **recontacter les patients** à la suite de découvertes de caractéristiques génétiques pouvant être responsables d'une affection justifiant des mesures de prévention ou de soins à leur bénéfice ou au bénéfice de leur famille, à l'exception des cas dans lesquels le patient s'y est opposé, **conformément à l'article L. 1130-5** du code de la santé publique ;



**Conséquences** : EDS du CAD

1- Application du référentiel sur l'Identité nationale de santé (INS)

2- Le retour vers la personne est un acte de soin → traçabilité et archivage ...

# Implication pour la constitution de l'EDS

---



# Information



Privilégier de multiples canaux d'informations

- Une information individuelle initiale sur la constitution de l'EDS
- Un portail de transparence
- De multiples canaux de communication : newsletter, affichage, réseaux sociaux ...

Portail de transparence : Permettre une information à plusieurs niveaux

- Pour chaque étude : avoir une notice d'information et présenter les résultats
- Permettre l'application des droits des personnes

[https://www.cnil.fr/sites/cnil/files/atoms/files/wp260\\_guidelines-transparence-fr.pdf](https://www.cnil.fr/sites/cnil/files/atoms/files/wp260_guidelines-transparence-fr.pdf)

---

# Cloisonnement des données

## Article 5.8 du référentiel EDS de la Cnil

Dans le cas où des données **directement identifiantes**, des **tables de correspondance**, des données **génétiques** ou des données de **suivi de localisation** sont versées dans l'entrepôt, celles-ci doivent être **stockées séparément** des données pseudonymisées, en utilisant les procédés décrits dans les exigences de sécurité SEC-LOG-4 à SEC-LOG-6 .

Nécessité d'avoir des **stockages distincts** au sein de l'EDS :

- Données identifiantes : nom, prénom, INS, adresse complète ...
- Données de santé
- Données de génétique
- Données du SNDS\*
- Données de localisation

\* **Ne peuvent être associées aux données identifiantes**



# Un Comité scientifique et éthique



Le comité est Scientifique ET Ethique

Article 3.2.3 :

3.2.3.1 : Seuls les projets ayant été examinés par cette instance peuvent avoir recours à l'entrepôt.

3.2.3.4 : Cette deuxième instance comprend notamment :

- au moins une personne impliquée dans l'éthique en santé ;
- une personne indépendante du responsable de traitement (par exemple : non salariée) ;
- des professionnels de santé et professionnels médico-sociaux ;
- des chercheurs ;
- un représentant des usagers ou d'une association de patients.

→ Indépendance du comité +++

---

## Des contraintes ...

---

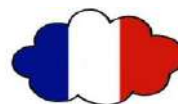


# Environnement sécurisé



Répondre aux exigences nationales et européennes

- Protection des données personnelles → RGPD
- Données de santé → EHDS
- Souveraineté → Attention aux fournisseurs extra-européens
- Sécurité → Sec\_num\_cloud, ISO:IEC 270001
- Sécurité des données de santé → HDS (certification hébergement de données de santé)



Et en plus encore ...

- SNDS : Homologation du système d'information au référentiel SNDS
- Soins : Cadre d'interopérabilité des systèmes d'information en santé
- IA : IA Act

# Environnement adapté à la volumétrie



1 BAM = 100 GO (WGS)

Cancer :

- WGS constitutionnel
- WGS somatique
- RNAseq

Maladies rares :

- Cas index
- 2 apparentés



Nécessité de recourir à des calculateurs de type HPC

- Collaboration CEA – CAD
- Vise un calculateur avec plusieurs dizaines de milliers de CPU +/- GPU

# Merci pour votre attention

---

Contact : Philippe Jean Bousquet

