

Le Collecteur Analyseur de Données (CAD) : Déploiement et accès aux données pour la recherche

Diane Gozlan¹, Christel Thauvin¹, Frédérique Nowak¹, Franck Lethimonnier¹, Yves Vandenbrouck², Abdelkader Amzert², Anne Jouvenceau².

¹ Equipe de Pilotage du Plan France Médecine Génomique, Inserm – Aviesan ; ² Collecteur Analyseur de Données (CAD)

INTRODUCTION

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

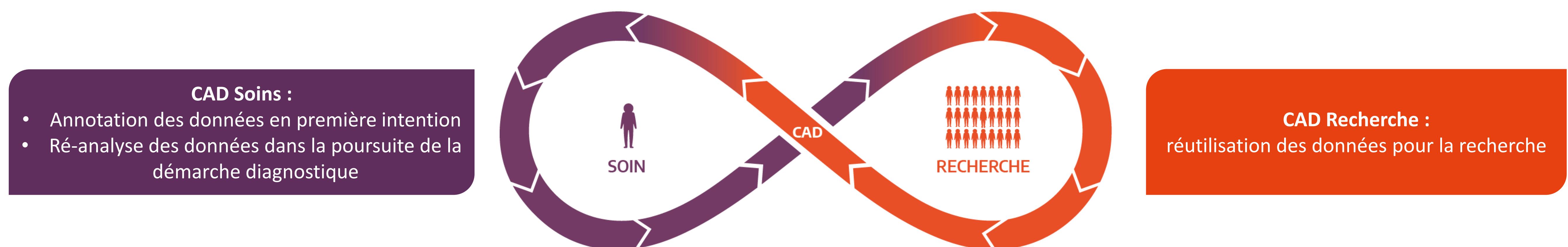
Il s'appuie sur 3 piliers principaux :

- un centre national de Collecte et d'Analyse des Données (CAD)
- un réseau de laboratoires de séquençage génomique (LBM-FMG pour Laboratoire de Biologie Médicale France Médecine Génomique)
- un Centre de Référence, d'Innovation, d'Expertise et de transfert (CReFiX).

La réutilisation des données issues du soin pour des projets de recherche en constitue un de ses axes majeurs, s'inscrivant dans une dynamique de science ouverte. L'ouverture des données a vocation à être la plus large possible, tout en assurant leur sécurité et en respectant un certain nombre de critères scientifiques et éthiques définis par le PFMG2025.

CAD : CONTINUUM SOIN-RECHERCHE

- **SOIN** : améliorer le diagnostic et la prise en charge des patients sur la base des connaissances générées par la recherche
- **RECHERCHE** : accélérer la recherche par la réutilisation des données PFMG2025

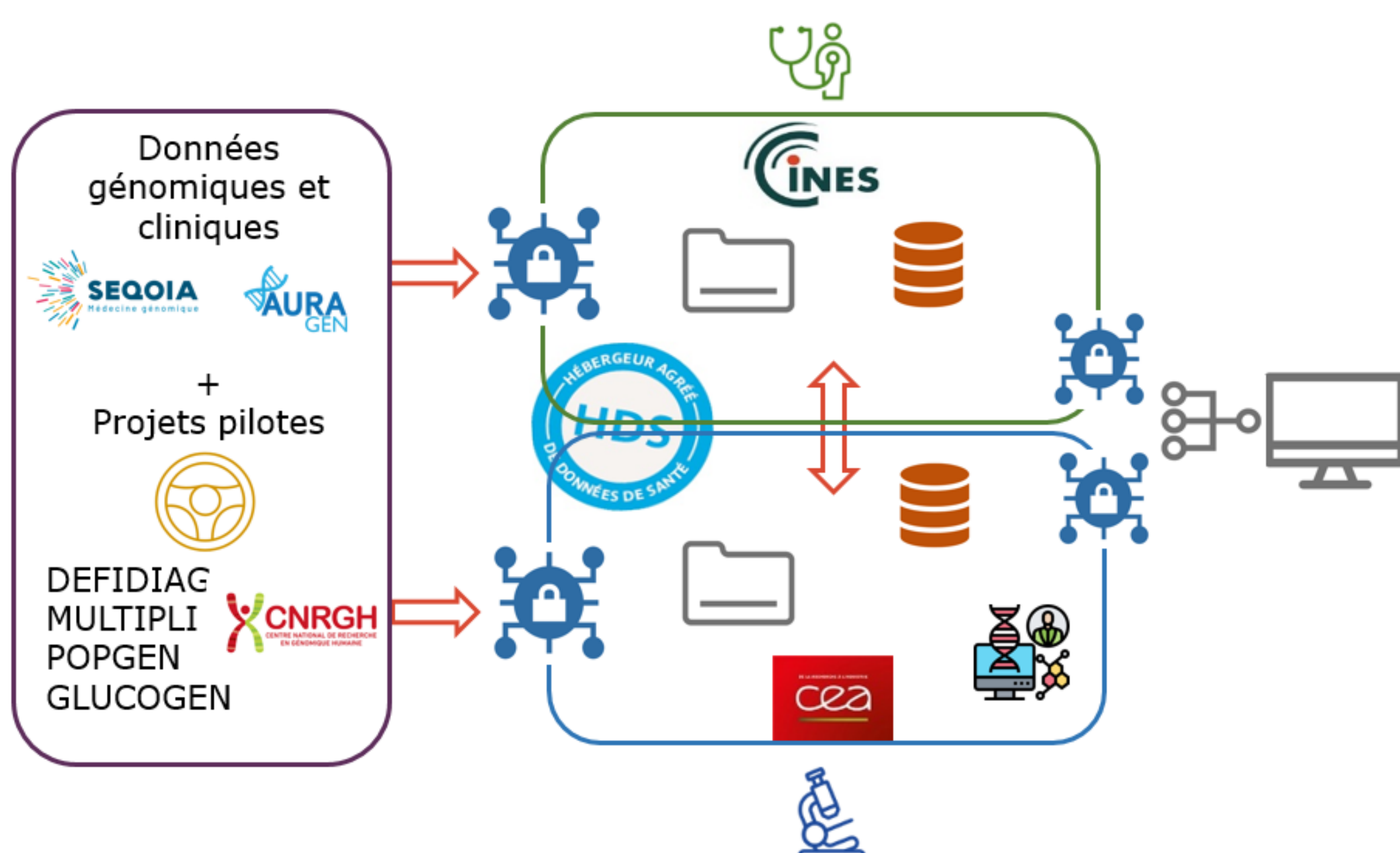


QU'EST-CE QUE LE CAD ?

Le CAD est une **infrastructure nationale** portée par la puissance publique, notamment **pour garantir la souveraineté des données, l'intérêt général et l'encadrement éthique** de l'utilisation des données. Il offrira à la communauté scientifique un **accès sécurisé au catalogue de données génomiques issues du PFMG**, un **éventail d'applications logicielles métiers**, des **facilités de stockage et de calcul intensif**

Le CAD sera hébergé au **Très Grand Centre de Calcul (TGCC) du CEA** en Ile-de-France et au **Centre Informatique National de L'Enseignement Supérieur (CINES)** à Montpellier. Le CAD sera alimenté en continu par les **LBM-FMG** (AURAGEN et SeqOIA à ce jour) et recevra également les données provenant **des 4 projets pilotes du PFMG2025** (MULTIPLI, DEFIDIAG, GLUCOGEN et POPGEN)

Le CAD s'appuie sur plusieurs groupes de travail et un **Comité Scientifique et Ethique (CSE)**. Ce comité pluridisciplinaire, a pour rôle de s'assurer que les projets de recherche ayant accès aux données respectent un certain nombre de **critères scientifiques et éthiques définis**



LES SERVICES DU CAD

- **Catalogue** : collecte/annotations des données génomiques du PFMG
- **Appariement** avec les données cliniques (si besoin)
- **Accès aux données** dans des « **bulles informatiques** » **sécurisées** au service des chercheurs
 - Mise en place d'espaces sécurisés par projet de recherche
 - transfert des données PFMG utile pour le projet de recherche
 - possibilité pour les chercheurs d'y apporter leurs propres données et outils d'analyse
- **Stockage et ressources en calcul** dimensionnées aux besoins du projet
- Portail et cellule d'**accompagnement** de projets / **formation**

PROCÉDURE D'ACCÈS AUX DONNÉES DU CAD POUR UN PROJET DE RECHERCHE

1. Demande d'accès aux données du CAD auprès du guichet d'accompagnement du PFMG2025
2. Dépôt du dossier de candidature au secrétariat du CSE
3. Évaluation de la conformité du projet de recherche aux critères d'accès aux données
4. Transmission de l'avis motivé (avis favorable, favorable sous réserve de maturation du projet, ou défavorable) du CSE au porteur de projet
5. Quand avis positif, échanges réguliers avec le porteur de projet et la cellule d'accompagnement du CAD pour définir et dimensionner l'environnement informatique et logiciel requis, et établir un calendrier prévisionnel
6. Signature d'un accord de collaboration et d'engagements du porteur
7. Ouverture d'un espace sécurisé dédié est créé au sein du CAD, afin d'accéder aux données du PFMG2025
8. Les prochaines dates de dépôt des demandes d'accès aux données et de réunion du CSE ont été définies

Date limite de remise des dossiers de demande d'accès au CAD	Date de réunion du CSE
9 février 2022	02 mars 2022
23 mars 2022	13 avril 2022
04 mai 2022	25 mai 2022
08 juin 2022	29 juin 2022

