

Janvier 2021

## Groupe de travail de préfiguration :

### Principes et modalités d'accès aux données du Plan France Médecine Génomique 2025 pour mener des projets de recherche

#### Contexte

Lancé en 2016 à la demande du Premier ministre et porté par Aviesan, le Plan France Médecine Génomique 2025 (PFMG2025) est un programme institutionnel national visant à intégrer de manière équitable sur le territoire la médecine génomique au parcours de soins. Celle-ci est d'abord introduite pour les patients atteints de cancers ou de maladies rares, avec une extension prévue aux maladies communes, en fonction de l'évolution des connaissances. L'ambition est que 50 000 patients en bénéficient chaque année. Le recours aux examens pangénomiques s'effectue selon une liste de pré-indications, un parcours de soins et une organisation territoriale définis pour répondre aux attendus du PMG2025. Les prélèvements sont envoyés aux laboratoires de séquençage (SEQOIA ou AURAGEN à ce jour) selon leur provenance géographique.

Au-delà de sa vocation sanitaire première, le PFMG2025 s'inscrit dans un continuum soin-recherche. La réutilisation des données issues du soin pour des projets de recherche constitue un des axes majeurs du Plan et a été anticipée dès sa phase de conception au travers de l'élaboration de modèles de consentements donnant la possibilité de consentir à la conservation des données obtenues dans le cadre du soin pour une utilisation ultérieure dans un but de recherche et du projet de création d'une infrastructure dédiée, le Collecteur Analyseur de Données (CAD) (cf annexe1).

Les données génomiques qui seront transmises au CAD sont constituées des fichiers contenant les variants génétiques (VCF), les fichiers contenant les données brutes (BAM) étant transmis au CAD au cas par cas, en fonction du besoin des projets de recherche. Les données cliniques seront constituées d'un set strictement nécessaire à l'interprétation clinico-biologique des résultats et à l'élaboration du compte-rendu de biologie médicale. Dans certains cas, des données de suivi des patients pourront aussi être collectées.

Certaines données transmises au CAD seront également issues de projets de recherche, en particulier les quatre projets pilotes du PFMG2025 pour lesquels le séquençage est effectué au CNRGH. MULTIPLI, DEFIDIAG, et GLUCOGEN sont des études cliniques. POPGEN est un projet en population générale associé à la cohorte Constances.

#### I- Définition des principes généraux guidant la réutilisation des données issues du PFMG2025 pour la recherche

La réutilisation pour la recherche des données générées dans le cadre du PFMG2025 s'inscrit dans une dynamique de science ouverte. L'ouverture des données doit être la plus large possible pour assurer une connaissance des maladies concernées la plus large possible.

Néanmoins, il s'agit de données de santé, de personnes malades ou non. Ces personnes ne les cèdent pas mais consentent à leur utilisation pour des projets de recherche, et ainsi à leur partage. D'autre part, ce sont des données génétiques, par nature identifiantes, non seulement pour la personne concernée mais aussi pour les membres de sa famille. Ces données sont par ailleurs principalement générées dans le cadre de parcours de soins, très divers, allant de prises en charge précoces pour le diagnostic de maladies rares, à des prises en charge de fin de vie pour des récidives de cancers. Dans ce contexte, l'utilisation des données pour la recherche n'est pas envisagée par les patients en première intention, mais ne constitue pour eux qu'une possibilité au-delà de leur parcours de soin.

De ce fait, la réutilisation de ces données ne peut être envisagée sans un engagement fort du PFMG2025 et des chercheurs envers les personnes ayant consenti au partage de leurs données pour la recherche.

Elle implique également une reconnaissance des chercheurs envers le programme institutionnel ayant conduit à la mise en place des infrastructures et des organisations permettant la génération et la réutilisation des données pour la recherche.

Dans ce contexte, la réutilisation des données issues du PFMG2025 pour la recherche doit être guidée par les principes suivants :

- 1 - Ouverture la plus large possible des données pour favoriser la recherche ;**
- 2 - Reconnaissance des personnes consentant à la réutilisation de leurs données comme étant une des parties prenantes de ce partage :**
  - Assurer la transparence sur l'utilisation des données partagées et les résultats qui en découlent ;
  - Associer ces personnes au partage des données ;
- 3 - Contribution du partage des données *in fine* au bénéfice des citoyens et au bien commun ;**
- 4 - Respect des droits fondamentaux des personnes consentant à la réutilisation de leurs données ;**
- 5 - Protection de la vie privée des personnes consentant à la réutilisation de leurs données :**
  - Assurer la sécurité des données ;
  - Ne partager que si c'est utile et seulement ce qui est indispensable ;
- 6 - Reconnaissance du programme institutionnel mis en place pour générer les données et permettre leur réutilisation pour la recherche.**

### **III- Modalités de mise en œuvre de ces principes**

Ces principes s'appliquent à tous les projets de recherche, qu'ils soient portés par un organisme public, le secteur privé ou des associations.

Ils se mettent en œuvre à plusieurs niveaux interdépendants les uns des autres :

- les personnes consentant à la réutilisation de leurs données (gestion de l'information, consentements, participation) ;
- l'infrastructure mise en œuvre pour le stockage et le partage des données ;
- les projets de recherche réutilisant les données (critères et modalités d'accès aux données).

La mise à disposition des données concernera dans un premier temps les données de soin. Les données de recherche (projets pilotes) seront disponibles à l'issue des projets, avec des circuits spécifiques d'accès aux données déclinés pour chacun d'entre eux.

### **Information, consentements et participation des patients**

Le PFMG2025 dispose d'ores et déjà de modèles de consentements pour les parcours de soin et les projets pilotes, donnant la possibilité aux personnes de consentir à la réutilisation de leurs données pour des projets de recherche. Il s'agit d'un consentement élargi, c'est-à-dire donné pour l'ensemble des projets de recherche.

Les personnes pour qui des données seront générées dans le cadre du PFMG2025 devront avoir accès à un service dématérialisé de gestion des consentements, leur donnant ainsi la possibilité effective de retirer leur consentement à tout moment. Ce retrait ne pourra toutefois pas concerner les projets de recherche déjà en cours ou terminés.

Ces personnes devront recevoir une information claire, compréhensible et facilement accessible sur leurs droits et la protection de leurs données. A cet effet, des notices d'information ont été élaborées dans le cadre du soin et pour les projets pilotes. De la même manière, les personnes devront également être informées des projets de recherche en cours et des résultats des projets de recherche.

Des représentants de personnes pour qui des données seront générées dans le cadre du PFMG2025 devront être impliquées dans la gouvernance de l'infrastructure de stockage et de mise à disposition des données.

### **Infrastructure de stockage et de mise à disposition des données :**

Le CAD devra garantir une sécurisation effective de la base, que ce soit pour les modalités de stockage des données ou les modalités d'accès à celles-ci.

Il mettra en place un service d'accompagnement méthodologique des chercheurs pour garantir la faisabilité des projets de recherche et leur qualité méthodologique (en mettant par exemple à disposition des outils d'analyse des données ou en dispensant une action de conseil).

En lien avec le chapitre précédent, il devra inclure dans sa gouvernance des représentants des personnes dont les données seront hébergées dans le CAD et pourront être réutilisées pour des projets de recherche. Il proposera un service dématérialisé de gestion des consentements.

Il devra respecter le choix des personnes quant au partage de leurs données pour la recherche.

### **Critères et modalités d'accès aux données pour mener des projets de recherche**

#### ***Critères et modalités d'accès aux données agrégées***

Les chercheurs disposeront d'un accès large et gratuit aux données agrégées, constituées par les résultats de l'analyse statistique des données génomiques de groupes d'individus, sous réserve que l'analyse les ait effectivement rendues anonymes et que ces données ne présentent plus de risques de ré-identification.

#### ***Critères et modalités d'accès aux données individuelles***

##### **❖ Phase temporaire d'accès aux données au travers du Pré-CAD**

L'accès aux données issues du PFMG2025 va se faire de manière progressive, avec une phase temporaire pendant laquelle il se fera au travers du Pré-CAD (cf annexe 1). Le nombre de projets de recherche pouvant accéder aux données sera alors nécessairement restreint pour des raisons techniques. Cela permettra, dans une logique de collaboration entre les porteurs de projets et le PFMG2025, de mettre en place un processus de co-maturation à la fois des projets de recherche et des services proposés par le CAD.

Un suivi des projets retenus et une évaluation des modalités d'accès aux données seront effectués, afin de s'assurer de la robustesse de la méthodologie de validation des projets et du respect des principes généraux définis pour la réutilisation des données.

L'ouverture aux industriels implique de définir les contreparties entre les acteurs du PFMG2025 et de s'accorder sur les modalités d'accès aux données, entre prestation de service et collaboration. Ce travail sera mené en parallèle de la phase temporaire, pendant laquelle il n'est donc pas prévu d'ouvrir l'accès aux industriels.

#### ❖ Critères devant être respectés par les projets de recherche

Les données du PFMG2025 sont ouvertes à tous les champs disciplinaires et à toutes les questions de recherche, sous réserve que les projets respectent les critères suivants :

- ✓ Contribution du projet au bien commun et pertinence des usages prévus :
  - Contribution à la connaissance et au progrès en santé ;
  - Intérêt et pertinence scientifique du projet ;
  - Pas de mésusage des données de génétique ;
- ✓ Faisabilité du projet à partir des données disponibles ;
- ✓ Caractère indispensable du set de données PFMG2025 pour le projet de recherche ;
- ✓ Caractère non sensible des données extraites du CAD (résultats de l'analyse) après la conduite du projet ;
- ✓ Conformité au RGPD et au cadre réglementaire (cf annexe 2) :
  - Pour les équipes françaises, la conformité au RGPD est traitée dans le cadre de la méthodologie de référence MR004 ou de l'autorisation de la CNIL. Afin de ne pas être redondant avec les évaluations menées par la CNIL ou l'organisme de rattachement du chercheur, il sera suffisant de vérifier que le porteur de projet s'est effectivement engagé dans ces démarches. L'accès aux données ne pourra être effectif qu'après la finalisation des démarches réglementaires ;
  - Une évaluation plus complète devra être menée pour les équipes étrangères, avec des modalités différentes selon qu'elles proviennent de pays où le RGPD s'applique ou non.

Ces critères s'appliquent à tous les projets de recherche, qu'ils soient portés par un organisme public, le secteur privé ou des associations.

Ces critères sont de deux ordres différents : si celui concernant la contribution au bien commun et la pertinence des usages prévus peut entraîner un refus définitif du projet, les autres dépendent du niveau de maturation du projet et peuvent être remplis dans un second temps. Pour atteindre le niveau de maturation suffisant, le projet pourra alors bénéficier d'un accompagnement du PFMG2025 dans le cadre d'une collaboration.

#### ❖ Comité Scientifique et Ethique

Un Comité Scientifique et Ethique (CSE) sera en charge de s'assurer que les projets ayant accès aux données respectent ces critères. Il présentera les caractéristiques suivantes :

- ✓ Gestion des liens d'intérêt :
  - déclaration des liens d'intérêt des membres du CSE et publication des DPI sur le site du PFMG2025 ;
  - gestion des liens d'intérêt avec les projets déposés. Les membres du CSE en lien avec un porteur de projet ne contribuent pas à l'avis émis sur ce projet ;

- ✓ Le CSE sera composé de membres permanents, avec la possibilité de solliciter un avis d'un expert extérieur si une compétence spécifique s'avère nécessaire pour l'étude d'un dossier ;
- ✓ Les membres du CSE seront nommés pour une durée de 3 ans ;
- ✓ Le CSE n'aura pas de statut d'IRB mais travaillera de manière coordonnée avec le CEEI de l'Inserm ;
- ✓ Le CSE sera composé d'une dizaine à une vingtaine de membres présentant une diversité de compétences :
  - représentants de patients pour les maladies rares et les cancers
  - représentant de la société civile et des citoyens/usagers du système de santé
  - représentants du service d'accompagnement du CAD (évaluation de la faisabilité du projet à partir des données disponibles et de la sensibilité des données extraites du CAD par les chercheurs après la conduite de leur projet)
  - juriste (évaluation de la conformité réglementaire)
  - chercheur en droit
  - expert en SHS en lien avec les questions de génétique
  - expert en éthique issu d'un CPP
  - expert en génétique des populations
  - expert en analyse bioinformatique des données
  - expert en informatique médicale
  - expert en sécurité informatique
  - cliniciens chercheurs dans le domaine de l'oncologie et des maladies rares, puis pour chacun des autres domaines d'applications du Plan FMG
  - représentants des ministères en charge de la santé et de la recherche

Les membres du CSE pourront être des francophones exerçant à l'étranger.

#### ❖ Procédure d'évaluation des critères devant être respectés par le projet

Suite à la procédure d'évaluation, le CSE pourra donner un avis favorable, défavorable ou favorable sous réserve de maturation du projet. En cas d'avis favorable, l'accord définitif sera donné par le CAD, afin par exemple de maîtriser le plan de charge du CAD.

La confidentialité des projets sera respectée tout au long de la procédure. En particulier, les membres du CSE signeront un engagement de confidentialité.

Un secrétariat sera mis en place pour gérer la procédure.

Les projets « standards » pourront être évalués selon une procédure accélérée. La notion de « projet standard » sera précisée par le CSE et pourra être revue annuellement.

La procédure complète, réservée aux projets soulevant des questions spécifiques, sera la suivante :

- ✓ Un dossier type sera à remplir par le porteur de projet ;
- ✓ Les projets seront reçus au fil de l'eau ;
- ✓ La fréquence de l'évaluation sera mensuelle. Tant que le nombre de projets reçus rendra cette procédure soutenable, les dossiers reçus seront répartis à l'ensemble des membres du CSE, avec la désignation de deux rapporteurs par dossier. Ils seront envoyés aux membres du CSE deux semaines avant la date de la réunion plénière mensuelle au cours de laquelle ils seront discutés ;
- ✓ Les avis du CSE seront motivés et transmis aux porteurs des projets ;
- ✓ Après l'avis positif du CSE, l'accord définitif sera donné par le CAD ;
- ✓ Après accord définitif, le porteur de projet signera un contrat type d'engagement.

Une attention particulière devra être portée au délai de réponse, afin de ne pas retarder les projets de recherche.

#### ❖ Engagements du porteur du projet

Le porteur du projet doit s'engager à :

- Rédiger un résumé du projet de recherche en langage compréhensible par tous qui sera mis en ligne sur le site du PFMG2025 ;
- Utiliser les données pour le seul projet décrit et ne pas les transmettre;
- À l'issue du projet, respecter le devenir prévu des résultats si ceux-ci constituent encore des données sensibles ;
- Rédiger un résumé des résultats de la recherche en langage accessible à tous qui sera mis en ligne sur le site PFMG2025 ;
- Publier les résultats dans des revues en accès ouvert ou déposer les publications dans l'archive ouverte pluridisciplinaire HAL ;
- Partager les résultats et enrichir les bases de données du CAD avec les résultats des projets (logique d'embargo à appliquer) pour les projets financés par la puissance publique (pour les projets financés par le secteur privé la négociation sur les contreparties intégrera le point) ;
- Respecter la Charte de Publication du PFMG2025 et en particulier :
  - remercier dans les publications les personnes qui ont partagé leurs données ;
  - remercier le PFMG2025, en tant que programme national ayant permis la génération et la réutilisation de ces données, et éventuellement les infrastructures associées (par exemple les laboratoires de séquençage du PFMG2025 et le CAD).

#### ❖ Suivi de l'activité de partage des données et de l'avancement des projets

Une cartographie des projets déposés sera effectuée (qui, quel sujet, quelles données, questions éthiques à analyser, etc...). Les résultats de ce suivi seront présentés au Comité Opérationnel du PFMG2025 (COMOP) afin de décider d'éventuels ajustements de la procédure. Une communication sera également effectuée sur le site internet du PFMG2025.

Dans une logique de collaboration et de co-construction, un suivi de l'avancement des projets sera mis en place au moins pendant la phase temporaire du pré-CAD, avec une présentation orale annuelle des projets. Cela permettra de s'assurer que les projets retenus se déroulent correctement et ainsi de rassurer les personnes sur le devenir de leurs données. Cela permettra aussi d'identifier des difficultés éventuelles de réutilisation des données et de mettre en place des solutions adaptées.

## Annexe 1 : Plateforme technique d'accès aux données

Parmi les outils développés spécifiquement dans le cadre du PFMG2025, le Collecteur Analyseur de Données (CAD) jouera un rôle central. Celui-ci a pour ambition de fournir une infrastructure de collecte des données génomiques et de mise à disposition de services ; en premier lieu sur le volet du soin pour aider les praticiens à interpréter ces données, ensuite sur le volet de la recherche pour offrir à la communauté de chercheurs la possibilité de mener des études sur les données collectées dans le cadre du soin. Le CAD offrira une capacité de stockage, une puissance de calcul intensif adaptée inédite à ce jour, un éventail de services informatiques et un accès à des bases de connaissances multiples.

Totalement lié aux autres outils du Plan, il sera alimenté en continu par les LBM de séquençage à très haut débit et recevra également les données provenant des projets pilotes. Le CAD sera hébergé dans deux des quatre centres nationaux de calcul intensif et aura accès à leurs moyens de calcul. Le CAD a été imaginé comme une infrastructure nationale unique portée par la puissance publique, notamment pour garantir la souveraineté des données, l'intérêt général et l'encadrement médical de l'utilisation des données. La gestion des données du CAD s'appuiera sur le principe FAIR (Findable, Accessible, Interoperable, Reusable). Le CAD sera hébergé dans des environnements HDS (hébergeur de données de santé) et l'accès aux données se fera dans des « bulles sécurisées ». Il reposera sur un principe d'accès aux données excluant la possibilité de copie ou d'extraction des données brutes. Il s'appuiera sur une base de données nationale pour le service au soin (analyse comparative des données d'un patient avec les données nationales équivalentes, outil de réanalyse complète d'une population de patients à finalité diagnostique, etc.) et une base de données pseudonymisées pour la recherche.

Toutes les données inscrites dans l'entrepôt de données pour la recherche sont associées à un consentement dédié à la réutilisation pour la recherche. A partir de cette base des données de recherche, des jeux de données pourront être mis à disposition dans des bulles sécurisées et des préanalyses pourront être réalisées pour produire des jeux de données anonymes. Assez vite, des solutions de cryptographies partielles ou totales seront mises en œuvre pour rendre les données anonymes. Les résultats de ces analyses devront être des données non identifiantes. Elles pourront être transférées ou rendues accessibles en tant que données ouvertes. Un accord de consortium établi entre les partenaires contribuant à la génération des données et l'institution de l'équipe utilisatrice des données encadrera les accès en termes notamment de tarification, publication, propriété intellectuelle, traçabilité, etc...

Dans l'attente de la mise en place des services intégrés prévus par le CAD, une solution temporaire permettant d'offrir uniquement le service de mise à disposition des données produites par le PFMG pour la recherche va être mise en œuvre. Préalable au CAD, le « Pré-CAD » pourra être opérationnel mi-2021 pour une durée de 18 mois. Il sera centré sur l'accès à des jeux de données dans un environnement sécurisé. Sur requête, les données nécessaires à un projet de recherche donné seront transférées au Pré-CAD par les laboratoires de séquençage du PFMG2025. Il n'y aura donc pas d'entrepôt de données national constitué. Les équipes de recherche pourront importer et exécuter leurs outils d'analyse. Un travail sera mené pour proposer quelques outils standards d'analyse mais il n'y aura pas de mise à disposition de bibliothèques d'outils. Les procédures de réanalyse seront réalisées par les chercheurs eux-mêmes.

## Annexe 2 : Critères réglementaires d'accès aux données PFMG2025

Il faut distinguer les critères réglementaires du PFMG2025 (cadre réglementaire du CAD, des LBM de séquençage à très haut débit, de la conformité des consentements, etc.) de ceux liés à la mise en œuvre de projets de recherche utilisant les données de santé hébergées dans le CAD.

Les projets de recherche réutilisant les données issues du PFMG2025 sont des recherches n'impliquant pas la personne humaine<sup>1</sup>. L'avis d'un Comité de Protection des Personnes (CPP) n'est pas requis pour les recherches n'impliquant pas la personne humaine. Si le projet de recherche répond à des critères très précis, il est possible de le réaliser dans le cadre d'une méthodologie de référence (MR-004 pour la réutilisation de données de santé existantes<sup>2</sup>). Les recherches en génétique peuvent entrer dans le cadre de la MR-004 si leur objet principal ou secondaire n'est pas l'identification ou la ré-identification des personnes par leurs caractéristiques génétiques. 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. Il doit effectuer une analyse d'impact relative à la protection des données, qui doit couvrir en particulier les risques sur les droits et libertés des personnes concernées. Il met en œuvre les mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté aux risques identifiés. La plateforme technique du CAD et le travail réalisé sur les consentements apporteront des réponses génériques à l'analyse d'impact. Si le projet rentre dans le cadre de la MR-004, il faut alors faire un engagement de conformité sur le site internet de la CNIL et enregistrer l'étude sur le répertoire public tenu par le HDH. L'ensemble des projets menés par un responsable de traitement dans le cadre de la MR-004 doit également figurer dans son registre interne des activités de traitement, conformément à l'article 30 du RGPD, et celui-ci peut être mis à la disposition de la CNIL sur demande.

Lorsque le projet n'est pas éligible à la MR-004, celui-ci doit être évalué par le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES) pour avis avant d'être soumis à la CNIL pour autorisation, le tout dans des délais encadrés.

Ces deux instances se prononceront sur :

- Le respect des droits et libertés des patients, en application du RGPD et de la Loi Informatique et Liberté (LIL) ;
- La sécurité des données.

---

<sup>1</sup> Article R1121-1 du Code de la Santé Publique : « Ne sont pas des recherches impliquant la personne humaine au sens du présent titre les recherches ayant une finalité d'intérêt public de recherche, d'étude ou d'évaluation dans le domaine de la santé conduites exclusivement à partir de l'exploitation de traitement de données à caractère personnel. »

<sup>2</sup><https://www.cnil.fr/fr/declaration/mr-004-recherches-nimpliquant-pas-la-personne-humaine-etudes-et-evaluations-dans-le>

# FRANCE MÉDECINE GÉNOMIQUE 2025

## Composition du groupe de travail de préfiguration

Frédérique Lesaulnier	Inserm	DPO
Catherine Bourgain	Inserm	directrice UMR Inserm, experte en génétique et sociologie, Comité éthique de l'Inserm
Flavie Mathieu	Inserm	Association de patients/ société civile
Marc Cuggia	Université de Rennes	Entrepot de données : projet Données Massives en Santé
Jacques Van Helden	Université d'Aix-Marseille	co directeur IFB, ELIXIR, expert en bioinformatique génétique
Daniel Zerbino	EMBL-EBI	EMBL-EBI à Cambriges, expert en bioinformatique /génétique
Ivo Gut	CNAG-CRG	Directeur du Centre d'Analyse du Génome, Barcelone, Président du Comité de sélection des plateformes FMG, Cancer Europe, membre de 1MG Europe
Jean-Louis Mandel	Personne qualifiée	expert maladies rares
Claude Ferec	Personne qualifiée	Personne qualifiée, expert maladies rares
Eric Guittet	Ministère de la recherche	Ministère de la recherche
Anne-Sophie Lapointe	Ministère de la santé	Ministère de la santé
Anne Cambon-Thomsen	Personne qualifiée	Personne qualifiée, expert éthique en génomique
Laura Richert	Université de Bordeaux	Plateforme Euclide/FCRIN/CIC, projet pilote Multipli
Nathalie Triclin	Alliance Maladies Rares	Présidente de l'Alliance Maladies Rares
Marie-France Mamzer	Université de Paris	Comité accès aux données entrepôt de données APHP

Estelle Lecointe- Artzner	ESMO	Présidente du CA de Sarcoma Patients Euro Net, European Patient Advocacy Group in the European Reference Network EURACAN
Christine Chomienne	Personne qualifiée	Mission Cancer Europe, B1MG Europe, experte Cancer

**Coordination Avisan :**

Diane Gozlan, Frédérique Nowak, Franck Lethimonnier