**Projet de recherche utilisant les données de l’EDS de la FESP**

Aide au remplissage du protocole

# Présentation de l’Équipe Projet

**Demandeur :**

* **Externe :** structures extérieures à la FESP (universités, organismes de recherche scientifique publics, associations, acteurs de santé, etc.)
* **Interne :** personnel de la FESP habilité, ayant une activité de recherche (chercheurs, post-doctorants, doctorants, statisticiens, internes de médecine, stagiaires...)

**Responsable de traitement (RT) :**

Le RT de données à caractère personnel est la personne, l’autorité publique, le service ou l’organisme qui détermine les finalités, les objectifs et les moyens de l’étude. Il endosse la responsabilité juridique du traitement et l’autorisation CNIL est établie à son attention.

Voici une liste d’indices permettant d’identifier le responsable de traitement :

* Il est à l’initiative du traitement.
* Il décide de l’objectif et des résultats du traitement.
* Il décide quelles données vont être collectées.
* Il décide quelle population va être concernée par le traitement.
* Il détermine des moyens et finalités du traitement.
* Il choisit les sous-traitants pour traiter les données pour son propre compte.
* Le RT doit être représenté par son représentant légal.

**Responsable(s) de mise en œuvre du projet (RMO) :**

Le **RMO du traitement** a la charge de réaliser l’étude ou une partie de l’étude. S’il n’est pas identique au RT, il est son sous-traitant.

**Délégué(e) à la Protection des Données (DPO) :**

Rôle du DPO : conseiller et accompagner l’organisme, contrôler l’effectivité des règles, être le point de contact de l’organisme sur les sujets RGPD, assurer la documentation des traitements de données, etc. cf. <https://www.cnil.fr/fr/le-guide-du-delegue-la-protection-des-donnees>

**Habilitation individuelle d’accès à l’EDS:**

L’habilitation individuelle donne accès à un espace de travail dédié et sécurisé au sein de l’EDS pour mener à bien le projet de recherche (PR).

NB : Les habilitations sont individuelles, nominatives et validées par la DP. Elles doivent être limitées en nombre, et toute personne habilitée doit avoir des compétences en gestion des données et analyse statistique.

**Méthodologie de référence (MR) :**

Les Méthodologies de Référence (MR) sont des procédures simplifiées d'accès aux données de santé. <https://www.cnil.fr/fr/recherches-dans-le-domaine-de-la-sante-la-cnil-adopte-de-nouvelles-mesures-de-simplification>

Voir la check-list MR004 : cf. HDH\_StarterKit\_Checklist conformité MR-004.docx

**Autorisation CNIL :**

<https://www.cnil.fr/fr/demande-dautorisation-dune-recherche-en-sante-les-informations-fournir-et-les-criteres-doctroi>

# 3.1. Justification de l’intérêt public

Voir la page <https://www.health-data-hub.fr/interet-public> pour plus d’informations.

# 3.2. Justification du respect de l’éthique

Voir la page <https://www.ofis-france.fr/espaces-thematiques/integrite-scientifique-ethique-de-la-recherche-deontologie/> pour plus d’informations.

# 4.2. Population d’étude

**Population d’étude :**

L’EDS comprend les données des régimes obligatoire (RO) et/ou complémentaire (RC) des **personnes majeures** gérées par MGEN et ayant donné leur accord (environ 3 millions), ci-après désignés « Assurés ».

|  |  |
| --- | --- |
|  | **Estimation des effectifs au 31/12/2024** |
| RO+RC | 1.8 millions |
| RO seul | 700 000 |
| RC seul | 400 000 |
| **Total** | **2.9 millions** |

**Périmètre géographique :**

La population est répartie sur l’ensemble du territoire français (y compris DOM-TOM). Il y a aussi des Assurés qui résident à l'étranger. Il faudra préciser dans le protocole si ces derniers doivent être exclus de la sélection.

**Information de localisation :**

Des informations géographiques pour les Assurés et les Personnels de Santé (PS) sont disponibles.

En résumé, du niveau le moins précis au plus précis :

* Département (Assurés et PS)
* Commune (Assurés uniquement)
* Adresse de résidence aléatoirisée (Assurés uniquement).   
  NB : une participation financière correspondant à la facturation d’un prestataire extérieur pour le géocodage des adresses pourra être demandée.

**Période de ciblage :**

Les données EDS sont disponibles rétrospectivement à partir du 01/01/2020.

Pour les études prospectives, les nouvelles données sont déversées mensuellement.

NB : À partir de novembre 2025, les codes affinés (CIP, NABM, UCD, CCAM, LPP) pour les données RO ne seront plus disponibles pour les nouvelles données déversées. Veiller à bien prendre en compte cette évolution des données dans votre protocole.

# 4.4. Liste des variables

NB : Les Demandeurs devront respecter le principe de minimisation dans le cadre de la sélection des données sélectionnées pour leur PR et justifier leur demande. Les données sélectionnées doivent être adéquates, pertinentes et limitées à ce qui est nécessaire au regarde des finalités pour lesquelles elles sont traitées. Cela constituera un critère d’admissibilité important.

**Description des catégories variables de l’EDS (voir le catalogue de données pour plus de précision) :**

|  |  |
| --- | --- |
| **Famille de données** | **Exemple de données** |
| **Données socio-administratives des Assurés** | Genre, date de naissance (mois et année), date de décès (mois et année), statut matrimonial, lieu de résidence (pays, département, code commune, adresses aléatoirisées), activité professionnelle (dont situation vis-à-vis de l’emploi, employeur, etc.), gestion (dont dates d’entrée et de sortie du RO, RC et de prévoyance, exonérations du ticket modérateur), assiette de la cotisation (estimation du revenu), statut ALD (oui/non, pas de détail du type d’affection) |
| **Codes affinés**  (CIP, NABM, UCD, CCAM, LPP) | Code affiné du médicament ou de l’acte, quantité, prix facturé, coefficient, date de délivrance ou d’exécution, base de remboursement… |
| **Prestations de santé**  (Prescripteur/exécutant/facturant) | Code acte, code de regroupement, dates de soins et de liquidations, médecin traitant, type d’établissement, dépense engagé, base de remboursement, taux RO, taux RC, montants remboursés, quantité, prix...  Numéros FINESS juridique et géographique si établissement, caractéristiques du professionnel de santé prescripteur, exécutant ou facturant (dont catégorie, spécialité, département, secteur de conventionnement) |
| **Données de prévoyance (non disponibles au 17/11/2025)** | Année de couverture, type de sinistre, offre (contrat/garantie) souscrite, nom de la prestation, montant mensuel de prestation versé, nombre mensuel de jours indemnisés et non indemnisés en cas d’arrêt maladie, année de liquidation |

# 4.5. Appariement de données

Il est possible d’apparier deux types de données :

**Données à caractère personnel :**

Préalablement au déversement dans l’EDS de la FESP, des données issues de PR, études et évaluations dans le domaine de la santé précédemment réalisés par des Demandeurs internes ou externes, et dont leur durée de conservation n’a pas expiré, pourront faire l’objet d’un appariement aux Données MGEN dans le respect de la Règlementation. Ces données peuvent provenir de diverses sources (notamment enquêtes, applications e-santé).

* Nécessitera une autorisation CNIL

**Données non individuelles et sans caractère personnel :**

Données agrégées à différents niveaux géographiques (région, département, commune, IRIS, etc.). Elles sont issues le plus souvent de bases de données en accès libre, généralement gérées par des ministères / agences publiques.

**Exemple non exhaustif**

|  |  |  |
| --- | --- | --- |
| **Source** | **Exemples de catégories de données concernées** | **Justification** |
| DREES  (<https://data.drees.solidarites-sante.gouv.fr/>) | Données sur la santé et le système de soins, le système de protection sociale, la retraite, le grand-âge et l’autonomie, les minima sociaux et la pauvreté, le handicap et l’invalidité, l’enfance, la famille et la jeunesse, etc. | Caractérisation de l’état de santé de la population générale, des établissements/PS), des dépenses/ remboursements de soins, des territoires, etc. |
| INSEE  (<https://www.insee.fr/fr/statistiques>) | Données sociodémographiques et économiques | Caractérisation de la population générale, des établissements/PS, des territoires, etc. |
| Plateforme ouverte des données publiques françaises | Données sur la population générale, les établissements de santé, les professionnels de santé, le système de protection sociale, les territoires, les médicaments, etc. | Caractérisation de la population générale, des établissements/PS, des dépenses/ remboursements de soins, des territoires, des médicaments, etc. |
| IGN  (<https://www.geoportail.gouv.fr/>) | Occupation du sol (bâti, infrastructures de transports, espaces verts, etc.) | Caractérisation de l’environnement géographique |
| GD4H  (<https://gd4h.ecologie.gouv.fr/catalogue>) | Expositions environnementales (pollution de l'air, bruit, pesticides, etc.) | Caractérisation de l’environnement géographique |

**Clé d’appariement :**

Cas des données à caractère personnel : exemple : numéro adhérent MGEN

Cas des données non individuelles et sans caractère personnel : exemple : code commune (pour les données INSEE)

# 4.7. Traitement et analyse des données

Logiciel mis à disposition (au 17/11/2025) : STATA 19.

Logiciels à l’étude : R, Python et QGis (<https://qgis.org/>).

NB : Il faudra contacter l’équipe opérationnelle de l’EDS ([eds.fesp@mgen](mailto:eds.fesp@mgen)) en amont, si besoin d’un logiciel autre que ceux listés, et justifier dans le protocole son intérêt au regard du PR.