Projet d'appariement du Répertoire statistique des individus et des logements au Système national des données de santé

La Direction de la recherche, des études, de l'évaluation et des statistiques (Drees), qui calcule des taux de recours et des consommations moyennes à partir du Système National des Données de Santé (SNDS) dans de nombreuses productions, est confrontée à l'absence de consensus quant au choix d'une population de référence[1] dans cette source administrative. Les travaux méthodologiques menés jusqu'à présent, tant à la Drees que dans le réseau des organismes publics produisant des statistiques sur le SNDS, n'ont pas permis de réconcilier les champs des résidents au sens de l'Institut national de la statistique et des études économiques (Insee) et des consommants de soins. Cela constitue une difficulté récurrente pour établir des statistiques sur le champ sanitaire portant sur les résidents, à partir du SNDS. Cette difficulté est accrue pour des statistiques infranationales ou restreintes à certaines catégories de la population, qui font l'objet de demandes croissantes de la part des usagers. La Drees souhaite améliorer la qualité de ses productions à partir du SNDS, et leur cohérence avec les statistiques démographiques et sociales de l'Insee et des autres services statistiques ministériels.

L'Insee s'apprête à proposer, à travers le programme de Répertoire statistique des individus et des logements (Résil), une offre de service d'appariement fondée sur un référentiel de personnes estimées résidentes en France à la date du 1er janvier. Cet univers de référence au 1er janvier 2025, construit à partir de la consolidation de nombreuses sources administratives, serait disponible en mars 2026.

L'objectif de ce projet d'appariement est d'utiliser le référentiel Résil pour identifier dans le SNDS les personnes résidentes en France au 1er janvier de l'année 2025, afin de fiabiliser les informations présentes dans le SNDS et d'établir des statistiques sur une population résidente de référence, cohérente avec celle de l'Insee, tant au niveau national qu'au niveau infra-national. Les statistiques ainsi produites seraient plus robustes aux aléas de gestion du SNDS et aux chocs affectant la demande de soins, comme ceux observés durant la crise sanitaire.

1. Intitulé du projet

Appariement Résil-SNDS.

2. Services producteurs

Le Lab innovation et évaluation en santé (Lab santé) de la Drees pilote un projet d'appariement des données de Résil, produit par l'Insee, et de celles du SNDS, produit par la Caisse nationale de l'Assurance Maladie (CNAM). La Drees, à l'initiative du projet, facilite les échanges entre la CNAM et l'Insee, deux institutions avec qui ses partenariats sont fréquents en tant que service statistique ministériel rattaché au ministère en charge de la santé et des solidarités. La Drees se chargera également de mettre en place les conditions juridiques permettant cet appariement, dans le cadre de la loi Informatique et Libertés et de l'article 193 de la loi de modernisation de notre système de santé. En particulier, le traitement devra faire l'objet d'une autorisation de la Commission nationale de l'informatique et des libertés (CNIL) après avis du Comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé (CESREES), dans la mesure où l'usage du SNDS apparié avec d'autres données excède l'étendue des accès permanents de la Drees et de l'Insee et ne

[1] C'est-à-dire une population au périmètre clairement défini, sans double compte, et comparable d'une année sur l'autre, sur laquelle calculer des indicateurs de consommation de soins, et à laquelle les rapporter pour calculer des taux de recours ou dépenses moyennes.

rentre dans le champ d'aucune méthodologie de référence homologuée et publiée par la CNIL à date (procédure d'accès aux données de santé simplifiée prenant la forme d'un engagement de conformité à la méthodologie de référence).

3. Services réalisant l'opération

Ce projet ne nécessite pas de collecte de nouvelles données, dans la mesure où il repose sur l'appariement de deux bases de données déjà existantes, constituées elles-mêmes par le chaînage de différents fichiers administratifs et médico-administratifs. L'Insee et la CNAM, producteurs respectivement du répertoire Résil et du SNDS, seront impliqués dans les opérations d'appariement comme détaillé ci-dessous. Il s'agira d'un appariement direct, s'appuyant sur le numéro d'inscription au répertoire national d'identification des personnes physiques (NIR, aussi appelé numéro de sécurité sociale).

Les rôles joués par l'Insee, la CNAM et la Drees sont les suivants :

- À partir de la population de référence 2025, l'**Insee** est responsable de l'envoi de l'ensemble des NIRs (ainsi que du sexe et de la date de naissance) des individus résidents ou susceptibles de l'être au 1er janvier 2025, à la CNAM, via la procédure « SAFE »[2] pour obtenir les « identifiants SNDS »[3] de ces individus. Ces informations (NIR, sexe, date de naissance exacte) ne sont pas restituées dans le répertoire Résil, dont l'identifiant est le CSNS (Code statistique non signifiant), mais sont présentes dans le répertoire national d'identification des personnes physiques (RNIPP). Le service d'enrichissement via Résil (créé au sein de l'Insee et réservé au service statistique public) peut effectuer un retour au NIR pour l'appariement avec les sources de données ne contenant pas le CSNS, ce qui est le cas des données du SNDS[4].
  L'Insee serait également responsable de l'envoi, à la Drees, d'informations issues du répertoire Résil, permettant de qualifier la résidence (commune et probabilité de résidence).
- La **CNAM** est responsable de la réception des « identifiants SNDS » obtenus à partir des NIRs transmis par l'Insee via la procédure « SAFE ». La CNAM extrait ainsi les données du SNDS nécessaires au projet, et les met à disposition de la Drees.
- La **Drees** est destinataire de la table « résidents 2025 » générée par la CNAM, et des données issues de Résil transmises par l'Insee, qui seront hébergées dans l'espace de travail de la Drees sur le portail de consultation du SNDS de la CNAM, ainsi que sur Big Data, le serveur interne sécurisé de la Drees[5].

Le circuit détaillant les étapes de la production de cet appariement est présenté en Annexe 2.

---

[2] Procédure obligatoire pour tout appariement direct avec le SNDS : l'appariement direct consiste donc à faire le lien entre des patients connus du demandeur (par leur NIR/date de naissance et sexe) et leurs données de consommation de soins du SNDS. La procédure passe par un cryptage irréversible des informations Numéro de sécurité sociale-Date de Naissance-Sexe pour transformer l'identifiant SNDS en clair en identifiant SNDS crypté.

[3] NIR bénéficiaire haché, et pseudo NIR (NIR assuré, date de naissance, sexe), chaînable au SNDS

[4] Le service mis en place par la Plateforme des données de santé, reposant sur le répertoire « miroir » du RNIPP de la CNAV (système national de gestion des identités, SNGI) pourrait constituer une alternative pour le retour au NIR à partir des traits d'identité disponibles dans Résil.

[5] Le serveur Big Data de la Drees est conforme à la Politique de sécurité des systèmes d'information de l'Etat et plus spécifiquement homologué conformément au référentiel de sécurité du SNDS, l'homologation la plus récente datant de 2024.

4. Nature de l'appariement demandé

Le présent projet porte sur l'appariement des données du répertoire Résil et de celles du SNDS.

4.1 Unité statistique

Les deux bases de données étant centrées sur les individus, la base Résil-SNDS aura également comme unité statistique les individus.

4.2 Champ de l'appariement

Le champ de l'appariement correspond à la combinaison du champ des deux bases de données :

- les individus Résil 2025 au sens large, à savoir les personnes résidant sur le territoire français au 1er janvier 2025 ou susceptibles d'y résider, ces dernières étant définies comme celles dont la probabilité de résidence en France au 1er janvier 2025 est élevée, mais pas suffisamment pour qu'elles soient intégrées avec certitude à la population de référence. Le champ couvert par Résil est intrinsèquement lié aux sources qui l'alimentent (Annexe 3), et correspond aux personnes résidant, travaillant, étudiant, ou recevant des prestations sociales en France ;
- les individus présents dans le référentiel des bénéficiaires du SNDS, à savoir ceux qui ont consommé au moins un soin ayant fait l'objet d'un remboursement par une caisse d'assurance maladie obligatoire ou ceux qui ont été hospitalisés, et qui sont vivants le 1er janvier 2025.

La bonne réalisation du présent projet nécessite de disposer de l'ensemble du champ des deux bases, et pas seulement de leur intersection, afin d'identifier et de caractériser trois populations : les résidents consommants de soins, les résidents non consommants de soins, ainsi que les consommants de soins non-résidents. Le premier objectif d'étude à partir de cet appariement, détaillé ci-après, est méthodologique : il s'agira de caractériser ces populations, de documenter les différences entre les deux bases de données appariées (comparaison des champs, du remplissage de variables communes), et d'en tirer les conclusions pour le choix de la population de référence sur laquelle fonder les statistiques produites par la Drees à partir du SNDS. Il s'agira également d'évaluer dans quelle mesure la population Insee peut être approchée à partir de critères observés dans le SNDS, apportant ainsi une évaluation complémentaire de la qualité du répertoire Résil (notamment en qualifiant les hypothèses faites sur les individus susceptibles de résider en France au regard de leur consommation de soin, en fonction des catégories d'âge et des territoires).

4.3 Champ géographique

Résil, du fait des sources qui l'alimentent, a pour champ géographique la France (métropole et DROM). Le SNDS est également disponible sur ce même champ. La base Résil-SNDS offrira la possibilité de faire des études sur l'ensemble de la France, ainsi qu'à des échelles géographiques fines (communes voire zonages infra-communaux [IRIS, QPV[6]]).

4.4 Taille de la population

La population des résidents en France est estimée à 68,6 millions au 1er janvier 2025, d'après l'Insee. La population Résil des individus résidents ou susceptibles de l'être devrait donc être un peu plus conséquente. La population des bénéficiaires dans le SNDS vivants au 1er janvier 2025 est généralement du même ordre de grandeur que la population Insee de l'année. La population totale comptabilisera

[6] Les IRIS (îlots regroupés pour l'information statistique) sont les plus petites unités spatiales statistiques, inférieures à l'échelon communal (les communes d'au moins 10 000 habitants et la plupart des communes de 5 000 habitants sont découpées en IRIS) ; QPV : quartiers prioritaires de la politique de la ville.

les individus présents dans les deux sources, ainsi que ceux présents dans seulement l'une ou l'autre de ces sources. La taille de la population finale reste inconnue du fait qu'elle contiendra des individus qui n'ont pas consommé en 2025 ou qui ont quitté le territoire.

4.5 Variables demandées

Les variables qui seront utilisées dans le cadre de ce projet sont les suivantes :

- Données de la base principale du SNDS
  L'ensemble des données du SNDS, c'est-à-dire les données relatives aux bénéficiaires (*e.g.* mois et année de naissance, sexe, commune de résidence, statut d'assuré), à leurs consommations de soins en ville et à l'hôpital (DCIR et PMSI), aux pathologies inférées (cartographie des pathologies), et le cas échéant aux dates et causes de décès (CépiDC). Ces données sont nécessaires pour couvrir les productions statistiques de la Drees fondées sur le SNDS (liste non exhaustive au point 6). La Drees dispose d'un accès permanent à ces données, avec un recul historique de vingt années.

- Données du répertoire Résil 2025
  Données démographiques sur la population des individus résidents, ou susceptibles de l'être, au 1er janvier 2025 (mois et année de naissance, sexe), lieu de résidence (issu du référentiel géographique Gaïa de l'Insee : commune, IRIS, QPV), probabilité de résidence sur le territoire (calculée par l'Insee en fonction des informations de présence issues de différentes sources alimentant le répertoire)[7], et variables indicatrices de la présence ou de l'absence dans chacune des sources.

L'Insee transmettra la commune de résidence des individus du répertoire Résil 2025, ce qui permettra de mieux spatialiser les consommations des résidents, et de le faire en cohérence avec les statistiques démographiques de l'Insee. Il ne s'agit pas d'une information nouvelle, absente du SNDS, mais de la fiabilisation d'une information existante. En effet, la commune de résidence[8] des bénéficiaires figure dans le SNDS mais n'est cohérente qu'autour de 85 % avec les données de résidence au sens fiscal[9], ce qui n'est pas surprenant : la commune présente dans le SNDS est celle inscrite dans les bases locales de l'assurance maladie, qui peut différer de celle connue par l'administration fiscale. La commune de résidence dans le SNDS est notoirement peu mise à jour et parfois mal renseignée : c'est une information qui provient des adresses déclarées par les assurés, qui ne notifient pas systématiquement d'éventuels changements de situation à leur caisse d'assurance, et dont la qualité de remplissage varie en fonction du régime d'affiliation à la sécurité sociale. L'information sur le lieu de résidence pourra le cas échéant être plus détaillée que celle présente dans le SNDS, selon les zonages d'études infra-communaux usuels du service statistique public (IRIS, QPV).

---

[7] La construction du répertoire Résil repose sur la méthode des « signes de vie », dont l'enjeu est de combiner les informations de présence issues de différentes sources pour estimer la probabilité que chaque individu réside effectivement sur le territoire, en accordant différents poids à chaque signe de vie, compte tenu des caractéristiques des sources dont ils sont issus. Cette probabilité n'est donc pas une donnée administrative, mais la résultante de travaux statistiques.

[8] Le code commune n'est renseigné que pour les communes de France. Son absence peut indiquer que l'individu ne réside pas sur le territoire français, ou que l'information est mal remplie, sans qu'il soit possible de distinguer l'un ou l'autre de ces cas.

[9] D'après l'analyse des données de l'EDP-Santé, un enrichissement de l'échantillon démographique permanent par les données du SNDS déjà exploité à la Drees.

Les autres informations issues de Résil (mois et année de naissance, sexe, probabilité de résidence sur le territoire, indicatrice pseudonymisée de la source privilégiée) seront mobilisées pour l'expertise de l'appariement, comme détaillé dans la rubrique portant sur la « Nature des travaux statistiques prévus ».

4.6 Limites méthodologiques de l'appariement

L'appariement ne pose pas de véritable défi technique car il sera directement effectué à partir des NIRs : présents sous forme pseudonymisée (par un algorithme de hachage) dans le SNDS, et présents dans les bases de production de l'Insee (mais non restitués dans le répertoire Résil).

Il demeure cependant certaines limites dont il faudra tenir compte pour évaluer la qualité de l'appariement :

- Des individus disposant d'identifiants d'attente, délivrés avant l'attribution d'un NIR, sont présents dans Résil d'une part, et dans le SNDS d'autre part (*e.g.* NIRs temporaires pour migrants ou assurés provisoires). Un décalage temporel est à prévoir dans l'identification de ces individus dans les différentes bases à chaîner. Cela pourrait compromettre le rattachement des consommations des résidents qui n'ont pas encore de NIR aux résidents au sens de l'Insee.
- Côté SNDS, de très rares NIRs (moins de 1 % en 2021) ne sont pas correctement renseignés, notamment dans le cas où un individu consomme des soins en étant ayant-droit d'une autre personne, sans que son propre NIR soit connu de la CNAM[10]. C'est une des raisons pour lesquelles certains individus peuvent apparaître plusieurs fois dans le référentiel des bénéficiaires avec des pseudo-NIR différents. Par ailleurs, il existe des NIRs fictifs en vue de garantir l'anonymat pour quelques prestations, comme les interruptions volontaires de grossesses. De ce fait, certaines consommations ne sont pas rattachées à un NIR bénéficiaire, ce qui peut impacter l'appariement si l'individu ne consomme pas de soins par ailleurs, et remet en question la pertinence de la population de référence apportée par ce projet d'appariement pour des études portant précisément sur ces prestations (Vilain et al., 2024).
- Côté Résil, l'inclusion ou non des NIRs dans la population de référence repose sur l'identification des individus du répertoire qui résident effectivement sur le territoire national. La méthode mise en œuvre pour cela est celle des « signes de vie » (ou de présence) qui mobilise, en complément des données d'état civil incontestables (naissances, décès), l'information relative à la présence des individus dans telle ou telle source administrative. Les signes de présence dans chacune des sources peuvent ensuite être pondérés en fonction de la qualité de l'identification de l'individu et de sa pertinence au regard des individus concernés (Lefebvre, 2024). Cette méthode est en cours de développement et repose sur certaines hypothèses sous-jacentes, notamment pour le calcul d'une probabilité individuelle de présence sur le territoire par la combinaison des signes de vie, ou encore pour le choix du seuil permettant de déterminer la présence ou non d'un individu en France en fonction de cette probabilité.

5. Historique du projet, place dans le dispositif statistique existant et apport à la connaissance du domaine par rapport à d'autres sources déjà disponibles

[10] Pour la plupart, les caisses de remboursement de l'assurance maladie transmettent les NIRs propres aux ayants-droits (ou NIRs individuels) ce qui permet de les retrouver, mais elles ne le font pas toutes.

Ce projet est nouveau dans le format proposé (appariement entre une population Insee de référence et la population des consommants de soins), notamment parce qu'il repose sur un répertoire encore en cours de construction (Résil), qui doit aboutir à une population de référence 2025 à horizon mars 2026. Il n'y a donc pas de bilan d'exécution de projets similaires. Il s'inscrit néanmoins dans la continuité d'une réflexion collective qui a émergé depuis plusieurs années face à l'absence de consensus quant au choix d'une population de référence à laquelle rapporter les données de santé pour produire des statistiques sur l'état de santé et le recours aux soins de la population.

La Drees, et de nombreux autres acteurs du domaine de la santé (agences, observatoires, chercheurs), calculent des taux de recours ou de non-recours, ainsi que des consommations ou dépenses moyennes à partir du SNDS dans le cadre de leurs productions. Le choix de la population de référence pour ces estimations, qui ne laisse pas présager de difficultés particulières, occasionne en pratique des différences parfois importantes. Un rapport sur le sujet du RedSiam (2021), un réseau national regroupant des experts dans le domaine du SNDS d'horizons variés (institutions, agences, centres hospitaliers universitaires ou encore structures privées), invite à repenser la notion même de dénominateur, à s'interroger sur l'impact du choix de ce dernier, ainsi que sur l'interprétation des résultats obtenus à partir du SNDS.

Plusieurs méthodes et définitions coexistent pour l'estimation de la population de référence, sans qu'un consensus se dégage sur celles à privilégier. Par exemple, il est usuel de calculer des taux intra-SNDS (parmi les personnes ayant consommés des soins) ou encore de rapporter un décompte de consommations de soins calculé dans le SNDS à la population résidente établie par l'Insee sur la base des enquêtes annuelles de recensement et des estimations localisées de population. Des travaux de réconciliation sont alors nécessaires pour comparer les populations Insee et SNDS, dont les champs ne sont pas les mêmes. En particulier, l'Insee recense la population résidant en France alors que le SNDS peut contenir les personnes ayant un numéro de sécurité sociale mais ne résidant pas ou plus en France (notamment des retraités), en situation irrégulière et prises en charge par l'aide médicale d'État, ou encore non affiliées à la sécurité sociale mais prises en charge au titre de conventions internationales (travailleurs détachés, étudiants en mobilité internationale, touristes, etc.). Les informations présentes dans le SNDS, notamment le code commune de résidence connu de l'Assurance maladie, ne sont pas suffisantes pour exclure les individus domiciliés à l'étranger et ainsi rapprocher les champs. Même en se restreignant aux bénéficiaires de l'assurance maladie obligatoire connus de l'Insee (*i.e.* dont l'identifiant n'est ni fictif ni provisoire), et aux bénéficiaires de l'AME, rien ne garantit que la population ainsi définie constitue un sous-ensemble de la population de référence de l'Insee et inclut de façon exhaustive l'ensemble de la population résidente ayant consommé des soins en France. Or c'est l'hypothèse qui est faite, de façon implicite ou explicite, pour de nombreuses statistiques produites à partir du SNDS (sur la prévalence de maladies, le recours aux soins ou les dépenses de santé notamment).

La définition de la population de référence s'est avérée particulièrement problématique au cours de l'épidémie de Covid-19, avec un sujet de cohérence entre les indicateurs de suivi de la crise sanitaire publiés par plusieurs acteurs. En particulier, une étude de la Drees (Geay et al., 2022) a mis en lumière des écarts parfois importants entre les estimations de couvertures vaccinales contre le Covid-19 en France publiées en *open data* par trois institutions (Santé publique France, CNAM, Drees), et s'est penchée sur les différences méthodologiques à l'origine de ces incohérences. Si les raisons sont multiples, elles tiennent en grande partie au choix de la population de référence à laquelle est rapportée la population vaccinée (population de consommants, ou de résidents, estimées à dates variables).

Par ailleurs, les méthodes existantes pour le choix du dénominateur, proposées notamment dans le cadre du RedSiam, ont été mises à mal par la crise sanitaire, en raison du recours massif et sans précédent de l'ensemble de la population au système de soins, dans le cadre des mesures de dépistage et de prévention. Jusqu'en 2020, la comparaison des populations des consommants (ayant consommé des soins au moins une fois dans l'année) et des résidents (au sens de l'Insee), par année, conduisait à des résultats cohérents, c'est-à-dire que la population des consommants était systématiquement inférieure à la population résidente. En revanche, à partir de 2021, la population des consommants a considérablement augmenté, un choc qui s'est maintenu en grande partie en 2022, avec une population supérieure de l'ordre de 2 % à celle des résidents. Le rapport s'est de nouveau inversé en 2023. La hausse de 2021 et 2022 s'explique par le fait qu'un nombre important de personnes sans pathologie, qui n'auraient pas consommé de soins en l'absence de la pandémie, ont réalisé des tests antigéniques ou ont été vaccinées contre le Covid-19 en 2021 et, dans une moindre mesure, en 2022. Par ailleurs, la comparaison des populations de consommants et de résidents par pyramide des âges, ainsi que par région, montre que les problèmes de cohérence sont antérieurs à la crise sanitaire, qui les a substantiellement aggravés : il existait déjà des tranches d'âge quinquennales ou des régions où la population des consommants dépassait celle des résidents, un phénomène qui s'est accentué avec la crise.

Au bilan, la définition d'une population de référence est un sujet d'autant plus prégnant que la crise sanitaire a mis à mal les travaux antérieurs de réconciliation des champs ou d'explication des différences, raisonnablement crédibles jusqu'alors, et constituant depuis une réelle difficulté pour établir des statistiques de référence sur le champ sanitaire à partir du SNDS. Cette complexité est accrue pour des statistiques infranationales et/ou restreintes à certaines catégories de la population, qui font l'objet de demandes croissantes de la part des usagers.

6. Objectifs généraux du projet et principaux thèmes que l'appariement permettra d'aborder

L'appariement de Résil et du SNDS a pour objectif de fiabiliser les indicateurs publiés par la Drees, en fournissant un socle commun pour le calcul des taux de recours et de non-recours à tout type de soin, ainsi que pour l'estimation des dépenses de santé moyennes. Résil apportera une population de référence, celle des résidents en France au 1er janvier 2025 au sens de l'Insee, dont la consommation de soins pourra être identifiée dans le SNDS. Les données de soins seront ainsi rapportées à une population cohérente avec celle de l'Insee, une méthode directe et transparente, rendant les statistiques plus robustes aux aléas de gestion du SNDS et aux chocs affectant la demande de soins, comme ceux observés durant la crise sanitaire.

Plusieurs productions clés de la Drees bénéficieraient de ce projet :

- La base RAC, qui est une base de données simplifiée des dépenses et des restes à charge après remboursement par l'assurance maladie obligatoire (RAC AMO), construite à partir du SNDS[11]. Il s'agit d'un produit phare de la Drees, dont sont issues diverses publications ainsi qu'un outil de datavisualisation, permettant d'approfondir la connaissance de la distribution des dépenses de santé et de la part restant à la charge des patients. La base actuelle ne permet de calculer des statistiques que sur la population des consommants de soins, mais pas sur la population globale. Disposer d'une population de référence permettrait notamment d'étudier le non-recours et de tenir compte des individus qui n'ont aucune dépense de santé.

[11] Dont le dernier millésime est 2021, mais 2023 est en préparation.

- L'APL, ou indicateur d'accessibilité potentielle localisée aux soins, qui fait partie des jeux de données ouverts de la Drees mis à jour annuellement (*data.drees)*. Cet indicateur, développé par la Drees et l'Irdes, vise à mesurer l'adéquation spatiale entre l'offre et la demande de soins de premier recours à un échelon géographique fin. Il mobilise pour cela les données du SNDS ainsi que les données de population de l'Insee à l'échelle communale. Comme discuté au point précédent, la méthode consistant à rapporter des données de consommation issues du SNDS à la population au sens de l'Insee présente des limites. La construction de l'APL gagnerait en précision et en pertinence si cet indicateur était construit à partir d'une population de référence cohérente avec celle utilisée pour estimer la consommation de soins.
- Les comptes nationaux de la santé (CNS), l'un des satellites de la comptabilité nationale de l'Insee, dont ils adoptent la méthodologie, qui retracent les dépenses et analysent de façon détaillée la consommation finale effective[12] de biens et de services de santé en France ainsi que les financements correspondants. Ses principaux enseignements sont présentés à l'automne, en présence du ministre en charge de la santé, lors de la Commission des comptes de la santé, qui fait partie du cycle préparatoire du projet de loi de financement de la Sécurité sociale. Les comptes de la santé alimentent également les comparaisons internationales et servent de base à la réponse de la France au *System of Health Accounts* (SHA), base de données internationales sur les dépenses de santé gérée par l'Organisation de coopération et de développement économiques (OCDE), Eurostat et l'Organisation mondiale de la santé (OMS). La dépense courante de santé au sens international (DCSi) comprend les dépenses de la consommation de soins et de biens médicaux[13] (CSBM), ainsi que les dépenses liées à la gouvernance du système de santé, à la prévention et aux soins de longue durée. La DCSi est définie sur le champ de la consommation des résidents sur le territoire national pour les données harmonisées au niveau international. Cependant, pour les données françaises, le périmètre retenu est le même que celui de la CSBM, c'est-à-dire le champ des consommations effectuées sur le territoire national (sans soustraire la consommation des non-résidents sur le territoire national, ni ajouter la consommation des résidents hors du territoire national), faute de données pour distinguer notamment la consommation de soins des résidents de celles des non-résidents. Des travaux à partir de l'appariement Résil-SNDS permettraient d'identifier les dépenses des résidents sur le territoire national et de coller plus précisément au cadre posé par le SHA, afin de favoriser les comparaisons internationales.

Plus globalement, de nombreuses publications de la Drees calculent des taux de recours, avec des méthodes variables, sans qu'un consensus émerge. Les données de consommation de soins sont généralement rapportées à celles des consommants ou de la population Insee. Disposer d'une population de référence permettrait d'harmoniser les publications de la Drees et d'en rendre les résultats plus robustes. On peut citer notamment des publications ou jeux de données ouverts (*data.drees)* mis à jour régulièrement, où figurent des taux de recours calculés sur la population Insee :

---

[12] La consommation finale effective recouvre l'ensemble des biens et des services utilisés effectivement quelle que soit la manière dont ils sont financés.

[13] La consommation de soins et de biens médicaux (CSBM) comprend les soins hospitaliers, les soins ambulatoires (médecins, dentistes, auxiliaires médicaux, laboratoires d'analyse, thermalisme), les transports sanitaires, les médicaments et les autres biens médicaux (optique, prothèses, petit matériel et pansements). Seules les dépenses qui concourent au traitement d'une perturbation provisoire de l'état de santé sont prises en compte. Ainsi, les dépenses de soins aux personnes handicapées et aux personnes âgées en institution sont exclues. La CSBM couvre le champ des consommations effectuées sur le territoire national par les assurés sociaux français et les personnes prises en charge au titre de l'aide médicale d'État (AME) ou des soins urgents (soins des personnes en situation irrégulière sur le territoire français).

le panorama annuel des établissements de santé (taux de recours à plusieurs disciplines hospitalières ; Cazenave-Lacroutz dir., 2025), la publication annuelle dans la collection *Etudes et résultats* (ER) sur les causes de décès en France (Godet et al., 2025), le jeu de données ouvert mis à jour annuellement sur les gestes auto-infligés, l'Atlas de santé mentale (panorama de l'offre de soins en santé mentale et du recours sur le territoire français), le rapport sur l'état de santé de la population en France (Bagein et al., 2022), le panorama de la complémentaire santé (description du champ d'intervention de la complémentaire santé et de la couverture de la population française ; Lapinte et Solotareff dir., 2024), etc.

7. Nature des travaux statistiques prévus

Les travaux statistiques prévus sont de plusieurs natures. Il s'agira d'expertiser les données et de juger de la possibilité d'en faire un usage statistique. Elles pourront être utilisées pour produire des indicateurs de santé fiables dans les domaines de compétence de la Drees. Cette expertise de l'appariement permettra également d'apporter des informations sur la qualité du répertoire susceptibles d'intéresser l'Insee.

Constitution d'une population de référence et production de statistiques sur cette base

Les données, une fois validées, pourront être utilisées pour fiabiliser les productions de la Drees. Dans un premier temps, différentes populations seront comparées, en croisant les champs des résidents Insee et des consommants SNDS. L'impact du choix de l'une ou l'autre des populations sur la valeur des indicateurs de recours aux soins et de consommation moyenne de soins sera documenté. Ces enseignements s'inscriront dans la lignée des travaux méthodologiques sur le sujet du « dénominateur », qu'ils viendront enrichir. Ces travaux exploratoires pourront faire l'objet d'une première publication méthodologique, avant d'être intégrés, si cela s'avère pertinent, aux travaux de production statistique, d'études et d'analyses de la Drees sur les thèmes mentionnés au point précédent.

Expertise des données du répertoire Résil

Résil est un produit de l'Insee en phase expérimentale et la mise en place d'une offre de service d'enrichissement à partir de ce répertoire est toujours en cours. Si le projet d'appariement Résil-SNDS aboutit, la Drees sera parmi les premiers utilisateurs de ce service. Ce projet pilote pourrait apporter des enseignements précieux pour la consolidation du répertoire.

Les données directement identifiantes liées à l'utilisation de la carte vitale avaient été envisagées pour la constitution de Résil, une option écartée lors de la concertation éthique préalable à sa réalisation (CNIS, 2022), le sujet étant jugé trop sensible. L'objet du présent projet n'est pas de revenir sur ce choix : les données du SNDS exploitées par la Drees sont pseudonymisées de manière irréversible et ne peuvent pas alimenter le répertoire Résil. Néanmoins, ce projet d'appariement, en aval de Résil et ne comprenant que des données pseudonymisées, peut apporter un éclairage statistique précieux sous la forme d'une validation externe : l'existence ou non d'une consommation de soins dans l'année est un « signe de vie » fort. L'analyse de l'appariement pourra apporter des enseignements pertinents quant à la qualité du répertoire Résil, en qualifiant les populations ou les territoires où la probabilité de résidence et la fréquence de consommation de soins s'éloignent significativement, ce qui pourrait permettre de confirmer ou d'infirmer certaines hypothèses faites par l'Insee dans le cadre de sa constitution, notamment au sujet des individus dont la présence sur le territoire ne fait pas consensus dans les différentes sources qui l'alimentent ou de la source à privilégier pour estimer la probabilité de présence sur le territoire.

8. Origine de la demande et utilisateurs potentiels des résultats

Ce projet d'appariement est une demande émanant de la Drees, et les données en résultant seront réservées à ses seuls usages. Cela étant, le constat de l'absence d'alternative convaincante pour améliorer la qualité des indicateurs de recours en santé et le besoin d'avoir une solution fiable sont partagés par une pluralité de producteurs de statistiques en santé.

Le répertoire Résil apparaît comme une solution innovante, en réponse au défi de définir une population de référence dans le SNDS, et une amélioration significative pour les statistiques en santé est attendue de cet appariement. L'expertise que la Drees développera lors de sa réalisation, et les travaux issus de son exploitation seront précieux pour la communauté scientifique.

L'Insee pourra par ailleurs être intéressé par les enseignements tirés de l'analyse de cet appariement quant à la qualité du répertoire Résil, toujours en phase de consolidation.

Ce projet est une opportunité pour la Drees de jouer un rôle d'harmonisation sur les statistiques relevant de son domaine de compétences, et pour l'Insee de participer à la coordination institutionnelle dans le champ de la production de statistiques et à la cohérence des statistiques françaises. La Drees fournira des efforts en faveur de la transparence des résultats, au moyen de ses publications, et du partage, autant que possible, des programmes ayant servi à élaborer ses études, conformément à sa démarche *d'open source*.

9. Calendrier prévisionnel

La production de l'appariement dépend d'un certain nombre d'étapes dont les délais sont variables. Sa réussite sera en effet subordonnée à celle de plusieurs jalons, chacun constituant des étapes structurellement longues (file d'attente) ou étant des procédures nouvelles, dans les deux cas sujettes à d'éventuels retards : disponibilité et qualité de la population 2025 à l'Insee, traitement de la demande à l'Insee au niveau d'un nouveau guichet d'appariement, autorisation CNIL après avis du CESREES (rédaction, échanges, modification éventuelle des contours du projet), traitement de la demande à la CNAM, extraction des données du SNDS, expertise de l'appariement…

Pour autant, l'objectif est d'aboutir **au plus tard début 2027 à la mise à disposition d'un appariement expertisé** aux statisticiens de la Drees pour espérer pouvoir produire dans des temps raisonnables des statistiques sur l'année 2025.

<u>Calendrier prévisionnel :</u>

- Échanges pour définir les contours du projet avec l'Insee et la CNAM : T3 2025
- Avis d'opportunité CNIS : T4 2025
- Rédaction, validation et dépôt de la demande d'autorisation CNIL : T4 2025
- Avis du CESREES : T1 2026
- Autorisation CNIL : T2 2026
- Envoi des identifiants, Insee > CNAM : T3 2026
- Envoi des données métiers, Insee > Drees : T3 2026
- Réception données SNDS, CNAM > Drees : au plus tard T4 2026
- Expertise de l'appariement : T1-T2 2027
- Production de statistiques à partir de l'appariement : à compter du T3 2027

10. Instances de concertation préalables

Le projet d'appariement Résil-SNDS, par son envergure et sa dimension novatrice, va être discuté au sein de plusieurs instances de concertation du CNIS. Il sera d'abord présenté lors de la réunion du bureau du CNIS du 8 octobre 2025, puis devant la commission démographie et affaires sociales du 20 novembre 2025, à l'issue de laquelle un avis d'opportunité sera sollicité.

Le projet d'appariement sera ensuite soumis à l'avis du CESREES, le comité éthique et scientifique pour les recherches, les études et les évaluations dans le domaine de la santé, chargé de rendre, avant l'autorisation de la CNIL, des avis sur les projets d'études nécessitant le recours à des données personnelles de santé. Le CESREES est une instance indépendante qui évalue chaque dossier au regard : de la finalité et la méthodologie de la recherche, de la nécessité du recours à des données de santé à caractère personnel, de la pertinence éthique, de la qualité scientifique du projet, ainsi que du caractère d'intérêt public qu'il présente, le cas échéant.

L'orientation, le pilotage et le suivi de ce projet sont, et seront, assurés par la sous-direction de l'Observatoire de la santé et de l'assurance maladie (OSAM) de la Drees, dont fait partie le Lab santé, sous l'encadrement de Benoît Ourliac, sous-directeur d'OSAM.

11. Moyens humains et financiers mobilisés pour le projet d'appariement

Moyens humains *:* une *data scientist* du Lab santé de la Drees, encadrée par le chef de bureau ou son adjointe, sera affectée à ce projet. Des discussions sont en cours avec l'Insee, qui pourrait souhaiter être associé à l'expertise de l'appariement, dans la mesure où les résultats qui en seront issus pourraient intéresser les responsables du programme Résil.

Moyens financiers : la tarification de l'Insee à la Drees est en cours de définition.

12. Procédure juridique envisagée

La réalisation de cet appariement reposera sur une autorisation de la CNIL après avis du CESREES. La CNIL veillera à ce que le projet de traitement respecte les dispositions pertinentes du RGPD et de la loi Informatique et Libertés, s'appuyant pour cela sur une demande d'autorisation détaillant les caractéristiques du traitement envisagé, tant sur ses aspects juridiques que techniques. Des éléments précis devront notamment être communiqués sur les mesures mises en œuvre afin de protéger la vie privée et le respect des droits des personnes concernées, ainsi que la sécurité et la confidentialité des données (y compris la gestion du risque de réidentification).

Annexe 1 : Bibliographie

PUBLICATIONS

- Adjerad, R., Courtejoie, N. (2020, novembre). Pour 1 % des patients, le reste à charge obligatoire après assurance maladie obligatoire dépasse 3 700 euros annuels. Drees, *Études et Résultats*, 1171.
- Bagein G., Costemalle G., Deroyon T., Hazo J.-B., Naouri D., Pesonel E., Vilain A. (2022, septembre). L'état de santé de la population en France. *Les dossiers de la Drees*, 102.
- Barlet, M., Coldefy, M., Collin, C., Lucas-Gabrielli, V. (2012, mars). L'Accessibilité potentielle localisée (APL) : une nouvelle mesure de l'accessibilité aux soins appliquée aux médecins généralistes libéraux. Drees, *Études et Résultats*, 795.
- Bithorel P.-L., Reduron V. (2023, mai). Dépenses de santé et restes à charge pour les patients : comment expliquer les disparités entre départements ? Drees, *Études et Résultats*, 1265.
- Cazenave-Lacroutz, A. (dir.) [2025, juillet]. Les établissements de santé en 2023 - Édition 2025. Drees, *Panoramas de la Drees-Santé*.
- CNIS (2022). Le projet de Répertoire Statistique des individus et des Logements – RESIL. Rapport du groupe de concertation du Cnis (2022) avec mandat et composition du groupe, rapport du groupe et ses recommandations.
- Didier L. et Lefebvre G. (dir.) [2024, novembre], « Les dépenses de santé en 2023 - Résultats des comptes de la santé - Édition 2024 », *Panoramas de la Drees-Santé*, Drees.
- Dubost C.-L., Leduc A. (2020, septembre). L'EDP - Santé : un appariement des données socio-économiques de l'échantillon démographique permanent au Système national des données de santé. *Les dossiers de la Drees*, 66.
- Geay C., Haugomat T., Khaoua H., Suarez Castillo M. (2022, octobre). Les taux de personnes vaccinées et non vaccinées contre le Covid-19 en France - Comparaison des indicateurs en open data et tentative d'estimation. *Drees Méthodes*, 5.
- Godet F., Costemalle V., Aubineau Y., Fouillet A. Coudin E. (2025, juillet). Causes de décès en France en 2023 : des disparités territoriales. Drees, *Études et Résultats*, 1342.
- Koumarianos H., Lefebvre O., Malherbe L. (2024, juin). Les appariements : finalités, pratiques et enjeux de qualité. Insee, *Courrier des statistiques*, 11.
- Lapinte, A., Pollak, C., Solotareff, R. (dir.) (2024, juillet). La complémentaire santé : acteurs, bénéficiaires, garanties – Édition 2024. Drees, *Panoramas de la Drees-Santé*.
- Lefebvre O. (2024, juin). Le Répertoire Statistique des Individus et des Logements (Résil) – Un nouvel univers de référence pour les statistiques démographiques et sociales. Insee, *Courrier des statistiques*, 11.
- Legendre, B. (2021, septembre). Les trois quarts des personnes les plus éloignées des professionnels de premier recours vivent dans des territoires ruraux. Drees, *Études et Résultats*, 1206.
- RedSiam (2021, mars). Choix de la population de référence dans l'estimation de taux de recours aux soins, de taux d'incidence et de prévalences à partir des données du SNDS.
- Vilain, A., Fresson, J., Lauden, C. (2024, septembre). La hausse des IVG réalisées hors établissement de santé se poursuit en 2023. Drees, *Études et Résultats*, 1311.

Datavisualisations et jeux de données ouverts

- Les dépenses de santé et les restes à charges
- Jeu de données ou sur les gestes auto-infligés
- Atlas de santé mentale

Annexe 2 : Circuit des données : appariement direct avec le référentiel des bénéficiaires du SNDS par l'utilisation du NIR

Étape 1 – Retour au NIR pour les individus de la population de référence 2025 et transmission du NIR à la CNAM

Le répertoire Résil contient les traits d'identité des individus (nom, prénoms, sexe, date et lieu de naissance) et l'Insee gère par ailleurs le RNIPP, qui permet d'effectuer un retour au NIR. Ces bases de données sont conservées dans un entrepôt de données sécurisé dont l'accès n'est possible qu'aux personnes spécialement habilitées.

Un numéro d'identification non signifiant (ID_projet) est attribué de façon aléatoire à chaque personne résidente (ou susceptible de l'être) au 1er janvier 2025.

L'Insee transmet à la CNAM la table de pseudonymisation [ID_projet ; NIR] constituée, pour toutes les personnes concernées, du NIR, du sexe, de la date de naissance ainsi que du numéro d'identification non signifiant (ID_projet) créé spécifiquement et différent de l'identifiant communément présent dans les données sources. L'Insee le dépose sur la plateforme SAFE suivant les modalités précisées par la CNAM[14].

Étape 2 – Identification des individus résidents dans le SNDS et constitution d'une table de correspondance mise à disposition sur le portail de consultation de la CNAM et transmise à la Drees

Une structure déconcentrée de la CNAM applique aux NIRs en clair figurant dans le fichier normé (généré à partir de la table de pseudonymisation [ID_projet ; NIR]) les mêmes algorithmes de hachage utilisés lors de la consolidation de la base principale du SNDS. À partir de l'identifiant SNDS (aussi appelé pseudo-NIR) ainsi obtenu, la CNAM cible les personnes concernées de l'étude dans le référentiel des bénéficiaires du SNDS.

Elle extrait ainsi les données du SNDS nécessaires au projet, avec un identifiant SNDS haché spécifique au projet (ID_Drees), et construit une table de correspondance [ID_projet ; ID_Drees]. Ces données sont ensuite transmises à la Drees pour être hébergées sur son serveur interne homologué conforme au référentiel de sécurité du SNDS.

La CNAM met aussi à disposition de la Drees une table de passage [ID_projet ; ID_SNDS] sur le portail de consultation du SNDS, au même emplacement que les données du SNDS auxquelles la Drees a accès du fait de ses accès permanents.

Étape 3 – Transmission de données sur la population des résidents de l'Insee à la Drees

L'Insee réalise une extraction du répertoire Résil, pour toutes les personnes résidentes (ou susceptibles de l'être) au 1er janvier 2025, contenant : le même numéro d'identification non signifiant que celui transmis à la CNAM (ID_projet), certaines variables communes au SNDS qu'il s'agirait de fiabiliser (mois et année de naissance, commune de résidence), la probabilité de résidence en France (calculée par l'Insee en fonction des informations de présence issues de différentes sources alimentant le répertoire), ainsi que des variables indicatrices de la présence ou de l'absence de l'individu dans chacune des sources. L'Insee procède à l'envoi sécurisé de ces données à la Drees qui les héberge sur

[14] De façon alternative, cette étape de retour au NIR pourrait être réalisée via le service mis en place par la Plateforme de données de santé. Dans ce cas, l'Insee transmettrait à ce service une table [ID_projet ; état civil, sexe, date de naissance] qui se chargerait de retrouver le NIR dans le SNGI pour transmettre de façon sécurisée une table [ID_projet ; NIR] à la CNAM. Le reste du circuit des données est inchangé.

son serveur interne sécurisé, au même emplacement que les données transmises par la CNAM à la fin de l'étape 2.

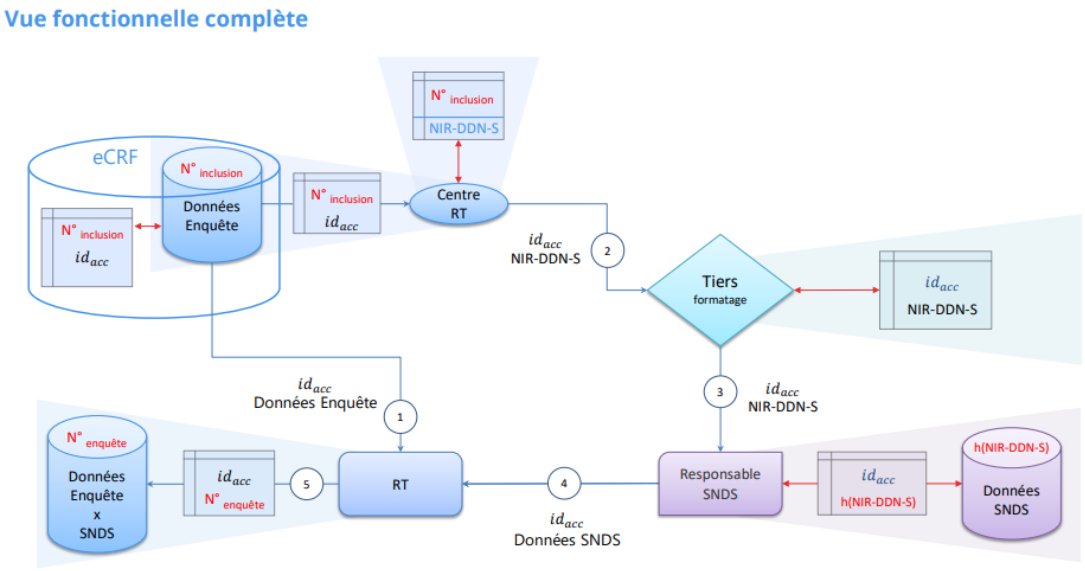
Étape 4 – Appariement des données par la Drees

À partir du numéro d'identification non signifiant (ID_projet), les tables de correspondances permettent l'appariement des données issues de Résil et du SNDS, sur le portail de consultation du SNDS de la CNAM d'une part, et dans un espace projet sécurisé de la plateforme de la Drees d'autre part, l'accès étant restreint dans les deux cas aux personnes habilitées. Les données appariées seront conservées pendant le nombre d'années autorisées pour la réalisation de l'étude et la publication des résultats, puis elles seront définitivement supprimées.

La Drees sera ainsi en capacité de calculer des statistiques 2025 fondées sur le SNDS et sur cette population de référence (taux, décomptes), soit sur le portail de consultation de la CNAM, soit sur sa propre infrastructure de calcul.

À noter que les chaînes de productions actuelles sont le plus souvent développées sur l'une ou l'autre des infrastructures, d'où la nécessité de disposer des données appariées sur chacune d'entre elles pour couvrir l'ensemble des productions. Les données mises à disposition sur le portail de la CNAM pourront être mobilisées dans le cadre des études Drees sur les APL ou pour les comptes nationaux de la santé notamment. Celles mises à disposition sur le serveur interne de la Drees permettront, quant à elles, d'alimenter la base RAC, qui y est hébergée de façon exclusive, et d'enrichir les études sur les taux de recours et dépenses moyennes qui en sont issues.

Circuit de la CNIL : synthèse de l'implémentation du circuit RT / centre unique / eCRF sans NIR



Annexe 3 : Sources utilisées dans le processus d'alimentation de Résil

La liste des sources utilisées pour construire et mettre à jour Résil été établie par un arrêté[15], publié au Journal Officiel le 5 janvier 2024.

Source	Justification
Répertoire national d'identification des personnes physiques (RNIPP)	• Mise à jour de la liste des individus avec les naissances et décès. • Mise à jour des données d'état civil.
Fichiers fiscaux : • **Fichier d'imposition des personnes (Fip)** • **Fichier permanent des occurrences de traitement des émissions (Pote)** • **Fichier de mise à jour des informations cadastrales (Majic)** • **R-Loc (Référentiel des Locaux)** • **Fichier « Gérer mes biens immobiliers » (GMBI)**	• Connaissance du parc de logements (Référentiel des Locaux, relais progressif des données du Cadastre). • Connaissance de leur usage (résidence principale ou secondaire) et de leurs occupants (via les données de GMBI pour l'occupant déclaré, Fip et Pote pour les autres occupants du logement). • Sont retraitées les données identifiantes et localisantes, à l'exclusion de toute information sur les bases fiscales, revenus fiscaux, impôts dus ou payés.
Déclaration Sociale Nominative (DSN)	La fréquence mensuelle de cette source permet une mise à jour fiable du répertoire d'individus (durée de présence sur le territoire, prise en compte de jeunes actifs en complément des données fiscales, prise en compte plus précoce des changements d'adresse).
Prélèvement à la source pour les revenus autres (PASRAU)	La fréquence mensuelle de cette source permet une mise à jour fiable du répertoire d'individus (durée de présence sur le territoire, prise en compte de bénéficiaires de revenus autres que salariaux, en complément des données fiscales).
Fichier de référence des allocataires Caf (Caisses d'allocations familiales)	Cette source apporte des éléments sur la composition des foyers d'allocataires, éléments indispensables à une information fiable sur les ménages (notamment leur composition).
Fichier d'allocataires de la MSA (Mutualité sociale agricole)	• Cette source apporte des éléments sur la composition des foyers d'allocataires, éléments indispensables à une information fiable sur les ménages (notamment leur composition). • Le champ de la MSA est complémentaire de celui des Caf.
Fichier annuel d'inscriptions dans l'enseignement supérieur	Ce fichier permet d'améliorer la couverture du répertoire sur les jeunes de 17 à 25 ans, parfois absents des sources fiscales ou localisés de manière ambiguë.
Répertoire des communautés constitué pour la réalisation du recensement de la population	Les communautés constituent la résidence principale d'environ 1,3 million de résidents sur le territoire. Il est donc essentiel de compléter le parc de logements ordinaires par celui des communautés. Le répertoire des communautés est un intrant du répertoire des logements de Résil.
Enquêtes statistiques de contrôle de la qualité du répertoire	L'Insee pourrait être amené à réaliser des enquêtes statistiques de contrôle de la couverture du répertoire, voire des enquêtes de complétude sur des territoires mal couverts par les sources administratives.

Source : Le Répertoire Statistique des Individus et des Logements (Résil) | Insee

[15] Arrêté du 5 janvier 2024 pris en application de l'article 2 du décret n° 2024-12 du 5 janvier 2024 portant création d'un traitement automatisé de données à caractère personnel dénommé « répertoire statistique des individus et des logements ».