VADÉMÉCUM

CIRCULATION DU NIR AUX FINS D'APPARIEMENT AVEC LE SNDS

Produit en collaboration avec le CASD

Version 1.0 – Juin 2024

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

1. Introduction

En raison des risques de réidentification, la CNIL est particulièrement vigilante sur les traitements prévoyant l'usage du numéro d'inscription au répertoire national d'identification des personnes physiques (NIR) comme identifiant pivot pour réaliser des appariements déterministes de données de santé avec le SNDS.

En appliquant un principe général de cloisonnement des identifiants utilisés, elle a publié en décembre 2020 un guide d'appariement avec le SNDS[1]. Ce guide est maintenant décliné sous forme de fiches pratiques, produites en collaboration avec le CASD, qui illustrent en détail des exemples concrets d'implémentation.

Le présent vadémécum accompagne les fiches pratiques[2].

Il en rassemble les principes communs, comme guide de lecture et aide-mémoire, et constitue également une grille d'analyse pour identifier les écarts et les axes d'améliorations d'un circuit existant et, si besoin, construire une variante personnalisée des circuits proposés.

Les principes sont repris ci-après sous la forme où ils apparaissent dans les fiches pratiques.

Présentation du CASD

Le Centre d'accès sécurisé aux données (CASD) est un groupement d'intérêt public (GIP) rassemblant l'État représenté par INSEE, le GENES, le CNRS, l'École polytechnique, HEC Paris et la Banque de France.

Il a été créé par arrêté interministériel du 29 décembre 2018.

Le GIP a pour objet principal d'organiser et de mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation. Il a également pour mission de valoriser la technologie développée pour sécuriser l'accès aux données dans le secteur public et dans le secteur privé.

[1] *Guide pratique : Modalités de circulation du NIR pour la recherche en santé aux fins d'appariement avec le SNDS* (PDF, 660 ko), CNIL, URL : https://www.cnil.fr/sites/cnil/files/atoms/files/guide_pratique_circuits_nir_recherche_en_sante.pdf

[2] « La CNIL publie de nouvelles fiches pratiques pour les circuits d'appariement avec le SNDS utilisant le NIR », CNIL, URL : https://www.cnil.fr/fr/fiches-pratiques-appariement-SNDS-NIR

Glossaire et abréviations

- **RT** : responsable de traitement
- **CU** : étude avec centre investigateur unique
- **MC** : étude multi-centres investigateurs
- **TC** : tiers centralisateur
- **eCRF** : cahier d'observation électronique
- **BCP** : base centrale pseudonymisée
- **SAFE** : plateforme sécurisée de transmission du NIR à la CNAM
- **NIR-DDN-S** : NIR, date et rang de naissance, sexe
- **h(NIR-DDN-S)** : NIR pseudonymisé, obtenu par « hachage cryptographique » de NIR-DDN-S
- **Id_{acc}** : « identifiant d'accrochage », identifiant technique temporaire non signifiant, permettant de dissocier le NIR et les données de santé lors des transferts entre acteurs

Couverture des cas d'usage par les fiches pratiques

	Données enquête en base locale		**Données enquête dans outil mutualisé**	
	NIR en base locale	***NIR dans outil mutualisé***	***NIR en base locale***	***NIR dans outil mutualisé***
Centre unique	RT – CU	Variante de MC – NIR chez TC	RT – CU – eCRF sans NIR	Variante de MC – eCRF avec NIR
Multi-centres	MC	MC – NIR chez TC	MC – eCRF sans NIR (ou MC – BCP)	MC – eCRF avec NIR

À noter :

- le cas « RT – CU » correspond à la fusion des cas « RT » et « CU » du guide de 2020, dans l'optique de maintenir une distinction et une séparation entre le centre investigateur (qui collecte les données) et le responsable de l'étude (qui va les exploiter), même quand les deux dépendent du même organisme.
- les circuits avec « reconstruction du NIR » (RN) n'ont pas été déclinés en fiche pour l'instant, dans l'attente du développement d'une solution automatisée permettant de solliciter les informations de la CNAV.

2. Principes communs

Généralités

- Le N° d'inclusion est présent sur des courriers, documents, échantillons et comptes rendus d'analyses biologiques médicales : il sera donc considéré comme un identifiant public.
- Les identifiants (d'accrochage, d'enquête, etc.) peuvent être générés par une fonction mathématique aléatoire, mais aussi par une fonction de hachage à clé secrète ; dans le second cas, c'est la clé secrète qui sera conservée au lieu de la table de correspondance.
- Le passage par un tiers centralisateur est utile quand le RT n'est pas autorisé à traiter le NIR des différents centres ; il peut aussi répondre à des contraintes organisationnelles (convention avec la CNAM, codes d'accès à SAFE et formatage).
- Sauf besoin dûment justifié (par ex. appariement récurrent, audit des données, alerte des participants), les tables de correspondance doivent être supprimées par les différents acteurs après envoi à l'acteur suivant du circuit, ou après finalisation de l'appariement dans le cas de la table entre les N° d'accrochage et les N° d'enquête.
- Toutes les communications doivent être sécurisés, notamment à l'aide de mesures d'authentification[3] et de chiffrement[4] conformes à l'état de l'art.

Pour les identifiants publics

- Le N° inclusion ne circule pas entre différents acteurs.
- Le NIR n'est stocké qu'à un seul endroit source et/ou éventuellement chez un tiers centralisateur, pour la durée nécessaire à l'inclusion des patients (le cas échéant) et à l'appariement des données avec le SNDS (date de fin du dernier appariement).
- Le NIR est stocké chiffré.
- Le NIR circule uniquement entre le détenteur d'origine (centre ou RT) et la CNAM, de manière chiffrée et pour être haché.
- Le NIR (ou le NIR haché) ne circule pas avec les données.
- La table de correspondance entre un identifiant et le NIR doit être isolée logiquement (dans une bulle ou par un chiffrement spécifique).

Les données du SNDS

- ne doivent pas être stockées avec un identifiant public (NIR ou N° d'inclusion) ;
- doivent être traitées dans une bulle sécurisée, distincte des outils de collecte des données de l'enquête (par ex. : eCRF) et conforme au référentiel de sécurité du SNDS ;
- doivent être accessibles uniquement par des personnes spécifiquement habilitées, distinctes des équipes assurant la collecte des données de l'enquête.

[3] Recommandation de la CNIL relative aux mots de passe et autres secrets partagés (PDF, 466 ko), CNIL, URL : https://www.cnil.fr/sites/cnil/files/atoms/files/deliberation-2022-100-du-21-juillet-2022_recommandation-aux-mots-de-passe.pdf

[4] Annexe B1 du référentiel général de sécurité (RGS), ANSSI, URL : https://cyber.gouv.fr/le-referentiel-general-de-securite-rgs

3. Principes appliqués aux différentes étapes des circuits

Saisie des données d'enquête

- Le NIR est cloisonné par rapport aux données d'enquête et au numéro d'inclusion (par exemple, le NIR est chiffré avec une clé spécifique).
- Chaque centre n'accède qu'aux données qui lui sont propres (sauf cas spécifique pour certaines BCP).

Génération des identifiants d'accrochage

- Utilisation d'identifiants techniques temporaires (« identifiants d'accrochage ») pour dissocier le NIR et les données de santé lors des transferts entre acteurs.
- Le N° d'accrochage est différent du N° d'inclusion afin de limiter les risques de réidentification croisée entre le numéro d'inclusion, le NIR et les données de santé (enquête et SNDS).

Extraction des données d'enquête / Envoi au RT

- Par principe, afin de limiter les risques de réidentification portant sur les données de l'enquête et sur les données du SNDS auxquelles elles seront appariées, le NIR et le numéro d'inclusion (ou tout autre identifiant utilisé pour le stockage des données) ne sont pas extraits des bases sources / transmis au responsable de traitement.
- Dès lors, seules les données strictement nécessaires à l'étude et à l'appariement (données d'enquête et identifiants d'accrochage) sont extraites des bases sources / transmises au RT.

Envoi des NIR au tiers

- Par principe, le numéro d'inclusion (ou tout autre identifiant utilisé pour le stockage des données) n'est pas transmis avec le NIR, ces deux identifiants posant par nature un risque élevé de réidentification. De même, le tiers n'a pas d'accès direct aux données stockées par les centres, et aucune autre donnée personnelle liée à l'enquête ne lui est transmise.

Envoi à la CNAM

- Envoi d'un fichier unique au bon format (d'où la nécessité d'un tiers centralisateur ou de formatage) via la procédure SAFE. Le fichier contient le NIR-DDN-S et le N° d'accrochage.

Envoi des données SNDS au RT

- Le pseudonyme h(NIR-DDN-S) ne sort pas de la CNAM et n'est visible par aucun autre acteur.

Appariement

- Les données SNDS ne peuvent être stockées que sur une plateforme SNDS nationale ou dans une bulle sécurisée, homologuée conforme au référentiel de sécurité du SNDS, et elles sont soumises à une habilitation d'accès spécifique.
- Aussitôt apparié, l'identifiant d'accrochage doit être remplacé par un nouvel identifiant d'enquête, seul identifiant visible par les chercheurs.