

Fiche pratique

CIRCULATION DU NIR
AUX FINS D'APPARIEMENT AVEC LE SNDS

CIRCUIT MULTI-CENTRES /
NIR CHEZ TIERS CENTRALISATEUR

Produit en collaboration avec le CASD

MC - NIR chez TC – Version 1.0 – Juin 2024

1. Introduction

La CNIL publie un ensemble de fiches pratiques, produites avec le CASD, présentant des exemples de circuits d'appariement avec le SNDS, en complément du guide pratique de la CNIL¹ publié en décembre 2020.

Les fiches présentent :

- des schémas fonctionnels détaillés pour chaque étape, dans le même formalisme que ceux du guide ;
- des schémas techniques orientés « tables de données », produits par le CASD.

Ces fiches illustrent en détail des exemples d'implémentation concrète des circuits du guide de 2020, lequel reste valide et se trouve ainsi précisé par les fiches.

Ces exemples respectent les principes issus du guide, qui ont été déclinés pour les fiches pratiques. Vous les trouverez rassemblés dans le document vademécum², comme aide-mémoire et guide de lecture.

La présente fiche concerne une étude multicentrique où chaque centre stocke localement ses propres données, sauf le NIR qui est centralisé dans un outil spécifique opéré par un tiers de confiance.

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 Vademécum : circulation du NIR aux fins d'appariement avec le SNDS (PDF, 328 ko), CNIL, URL :

https://www.cnil.fr/sites/cnil/files/2024-06/circuits_nir_vademecum.pdf

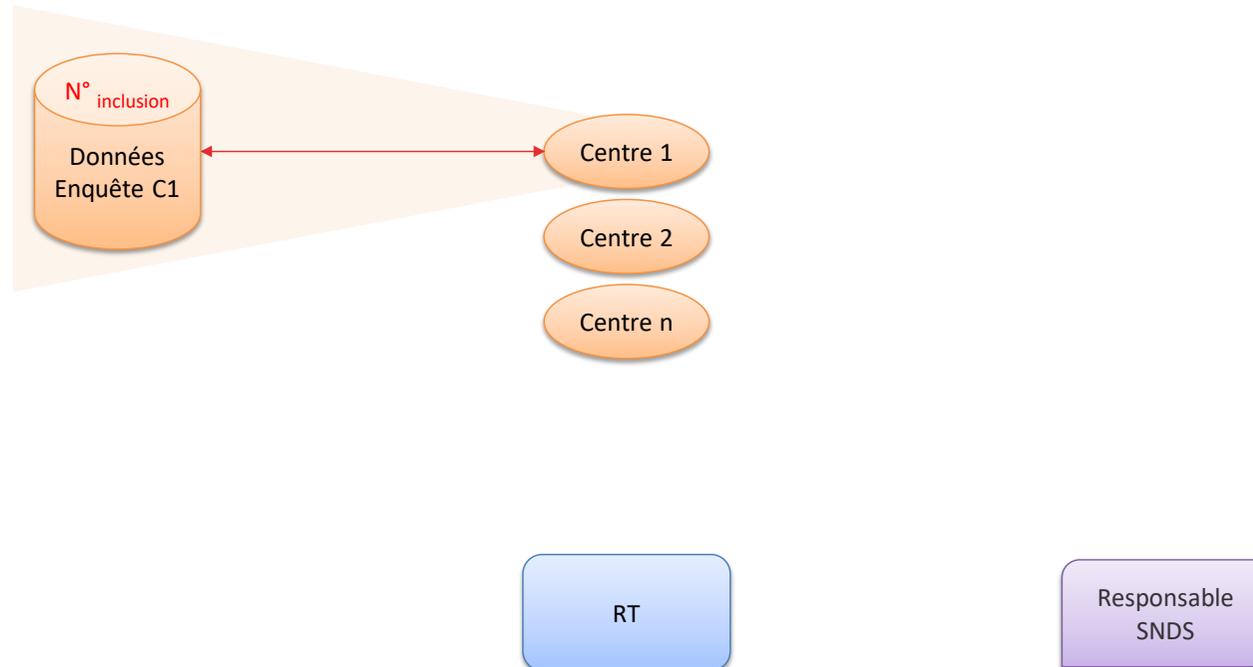
2. Implémentation détaillée du circuit Multi-centres / NIR chez Tiers centralisateur

Saisie des données d'enquête

Chaque centre saisit ses données dans une base locale ne contenant pas le NIR

- La sécurité est assurée localement selon les règles fixées par le centre et selon les consignes données par le RT.

- Afin de limiter les risques de réidentification, le NIR n'est pas stocké localement dans les centres mais saisi directement chez le tiers centralisateur (cf. étape 2).

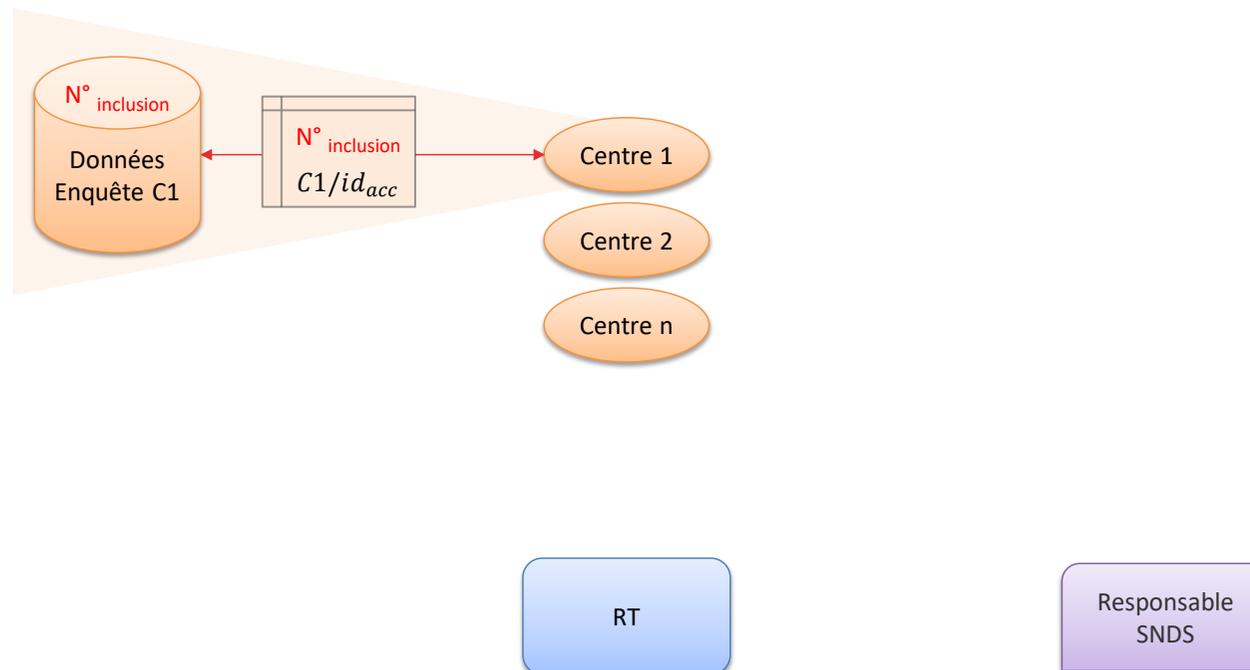


Etape 0 – Génération des identifiants d'accrochage

Chaque centre génère une table de correspondance entre le numéro d'inclusion des participants et un identifiant d'accrochage aléatoire et non significatif (C1/id_{acc} pour le premier centre, etc.)

- L'utilisation d'identifiants techniques temporaires (« identifiants d'accrochage ») permet de dissocier le NIR et les données de santé lors des transferts entre acteurs.
- Par principe, **ces numéros sont non significatifs et différents du numéro d'inclusion de la personne dans l'étude**, afin de limiter les risques de réidentification croisée entre le numéro d'inclusion, le NIR et les données de santé (enquête et SNDS).

- Si le besoin est dûment justifié (par ex. transmission récurrente, audit des données, alerte des participants), la table de correspondance entre le numéro d'inclusion et l'identifiant d'accrochage peut être conservée par chaque centre, de manière sécurisée.
- L'identifiant d'accrochage peut être généré 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.

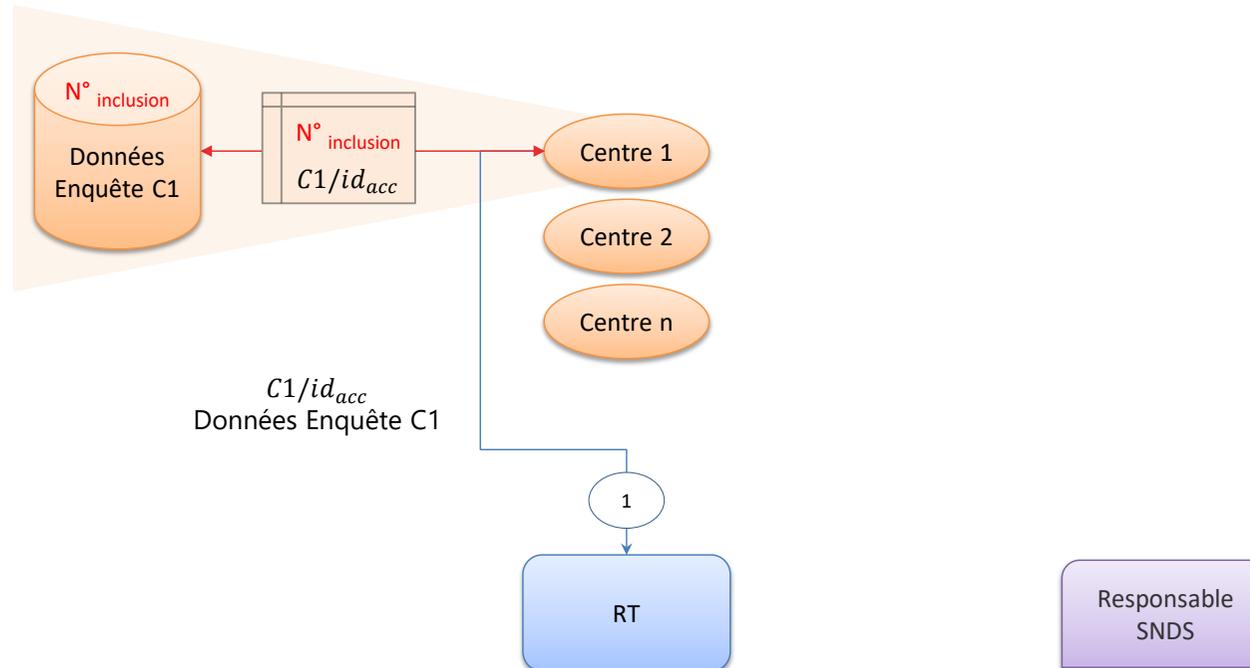


Etape 1 – Envoi des données d'enquête au RT

Chaque centre envoie au responsable de traitement les données d'enquête associées aux identifiants d'accrochage

- 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 ne sont pas transmis au responsable de traitement.**
- NB : les risques de réidentification sont à considérer pour la transmission et pour le stockage des données.

- Dès lors, **le RT n'a pas d'accès direct aux données stockées dans les centres** : il recevra de la part des centres uniquement les données strictement nécessaires (données d'enquête et identifiants d'accrochage).

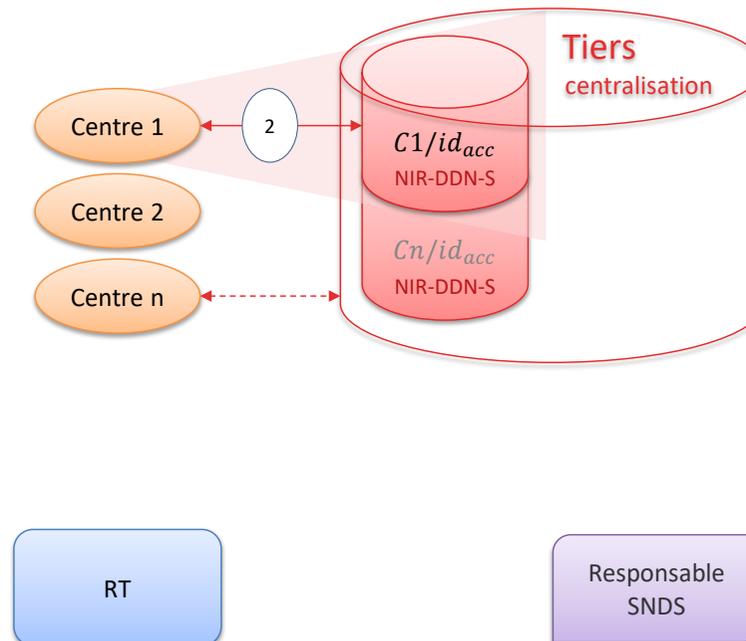


Etape 2 – Saisie des NIR chez le tiers centralisateur

Chaque centre saisit directement chez le tiers centralisateur les [NIR - Date de Naissance - Sexe] associés aux identifiants d'accrochage de ses participants

- **Les centres ne stockent pas localement les NIR.**
- Les [NIR - Date de Naissance - Sexe] sont saisis par chaque centre directement dans un outil opéré par le tiers centralisateur, avec les identifiants d'accrochage correspondants, comme seules données strictement nécessaires au circuit d'appariement.

- En effet, par principe, **le numéro d'inclusion n'est pas saisi 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 dans les centres, et aucune autre donnée personnelle liée à l'enquête ne lui est transmise.
- En raison des risques liés aux accès à distance via Internet et à la présence des données identifiantes [NIR - Date de Naissance - Sexe] dans l'outil opéré par le tiers, **sa base de données est chiffrée et une authentification forte est exigée à chaque connexion.**



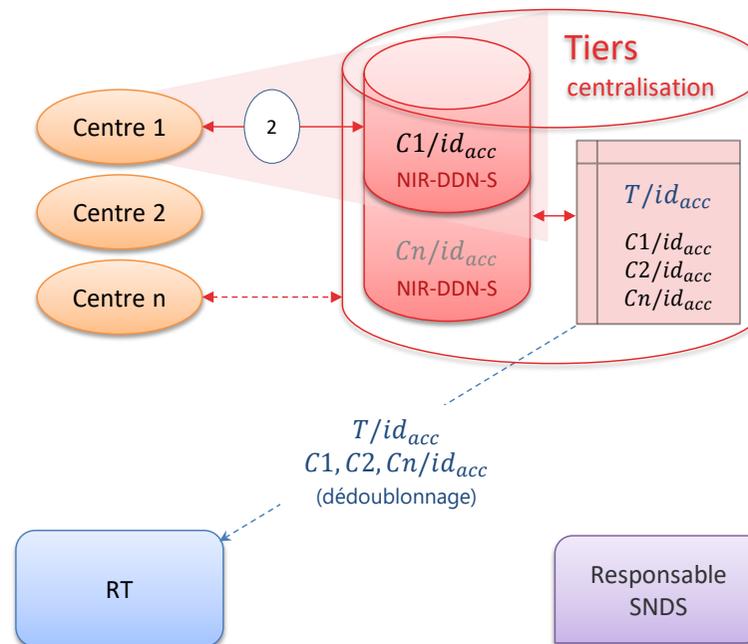
Etape 2bis – Dédoublonnage et envoi au RT

Si nécessaire, il est procédé au dédoublonnage des participants ayant le même [NIR - Date de Naissance - Sexe]

- Un participant qui serait suivi par plusieurs centres aurait plusieurs numéros d'inclusion et identifiants d'accrochage associés (par ex. $C1/id_{acc}$ et $C3/id_{acc}$).
- S'il est nécessaire de chaîner les données issues de plusieurs centres sur leurs participants communs, le tiers centralisateur identifie les lignes

avec le même [NIR - Date de Naissance - Sexe] et leur attribue un nouvel identifiant d'accrochage unique T/id_{acc} .

- Le tiers centralisateur transmet alors au responsable de traitement la table de correspondance entre les identifiants Cn/id_{acc} et T/id_{acc} afin de permettre *in fine* d'apparier les données des centres avec celles extraites du SNDS.
- S'il n'y a pas besoin de dédoublonnage, T/id_{acc} peut reprendre les Cn/id_{acc} .

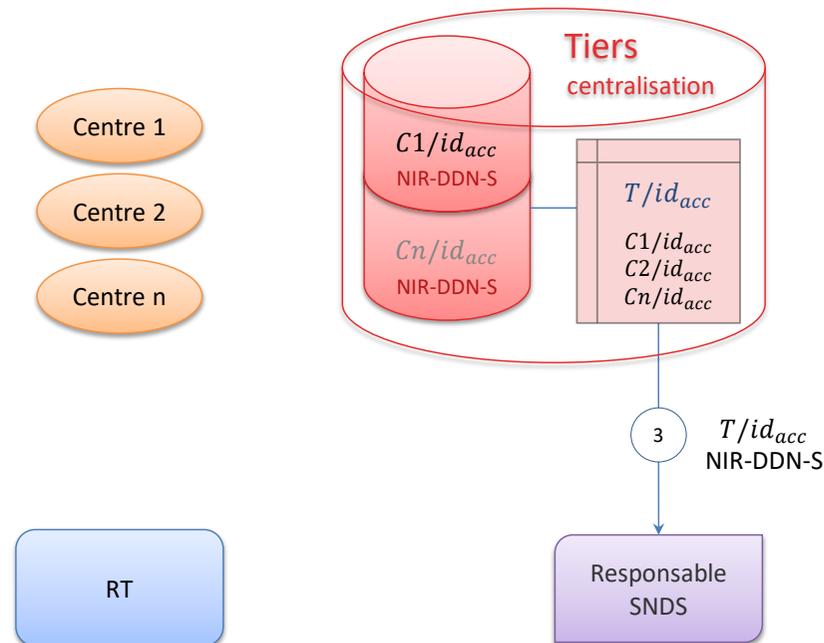


Etape 3 – Envoi des NIR au responsable SNDS

Le tiers centralisateur envoie au responsable de la base SNDS les (nouveaux) identifiants d'accrochage accompagnés des [NIR - Date de Naissance - Sexe]

- Cet envoi doit se faire à l'aide du téléservice « SAFE » de la CNAM, dans un fichier unique et au format adéquat.

- Si le besoin est dûment justifié (par ex. transmission récurrente, audit des données, alerte des participants), la table de correspondance entre les identifiants d'accrochage Cn/id_{acc} et T/id_{acc} peut être conservée par le tiers centralisateur, de manière sécurisée.

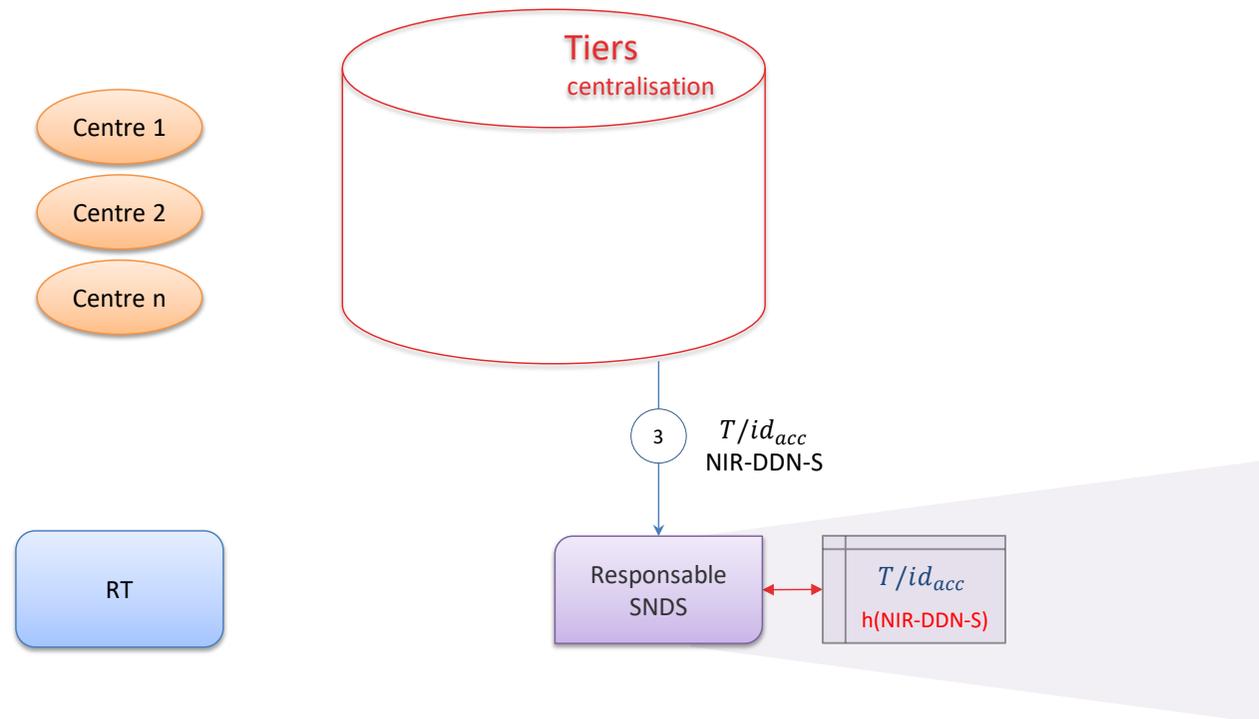


Etape 3bis – Hachage des NIR en entrée

Dès réception, le responsable de la base SNDS procède au « hachage » du triplet [NIR + Date de Naissance + Sexe] pour générer l'identifiant interne du SNDS : $h(\text{NIR-DDN-S})$

- Le hachage désigne ici un calcul cryptographique produisant une pseudonymisation irréversible.

- Dans le cas du SNDS, le NIR est pseudonymisé par plusieurs étapes de hachage successives.
- Par principe, le triplet [NIR + Date de Naissance + Sexe] est remplacé par $h(\text{NIR-DDN-S})$ dès réception des données et il n'est pas conservé, afin de limiter les risques de réidentification des données du SNDS.

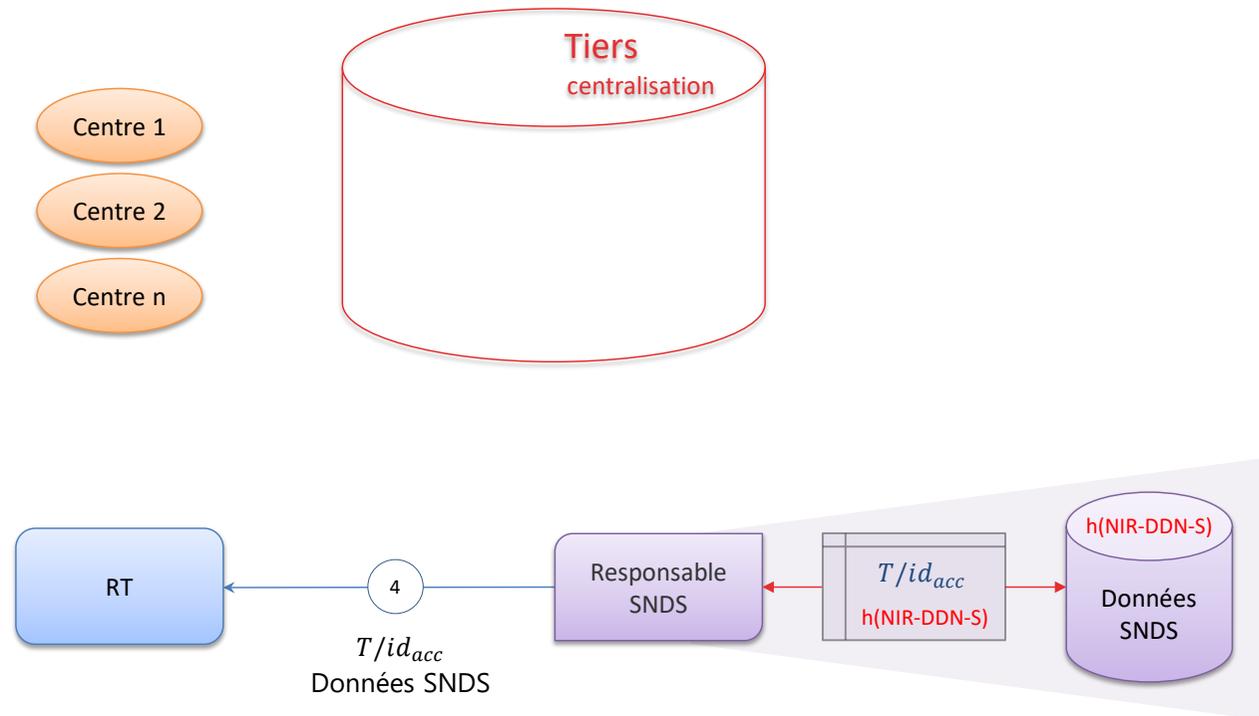


Etape 4 – Extraction et envoi des données SNDS au RT

Le responsable de la base SNDS extrait les données du SNDS correspondant aux $h(\text{NIR-DDN-S})$ des participants

Il transmet les données extraites au responsable de traitement, avec l'identifiant d'accrochage reçu du tiers centralisateur

- Afin de limiter les risques de réidentification des données du SNDS, son identifiant interne $h(\text{NIR-DDN-S})$ n'est jamais extrait : **seul l'identifiant d'accrochage est présent avec les données extraites du SNDS**



Etape 5 – Appariement des données d'enquête avec les données du SNDS, et génération d'un nouvel identifiant pour la base appariée

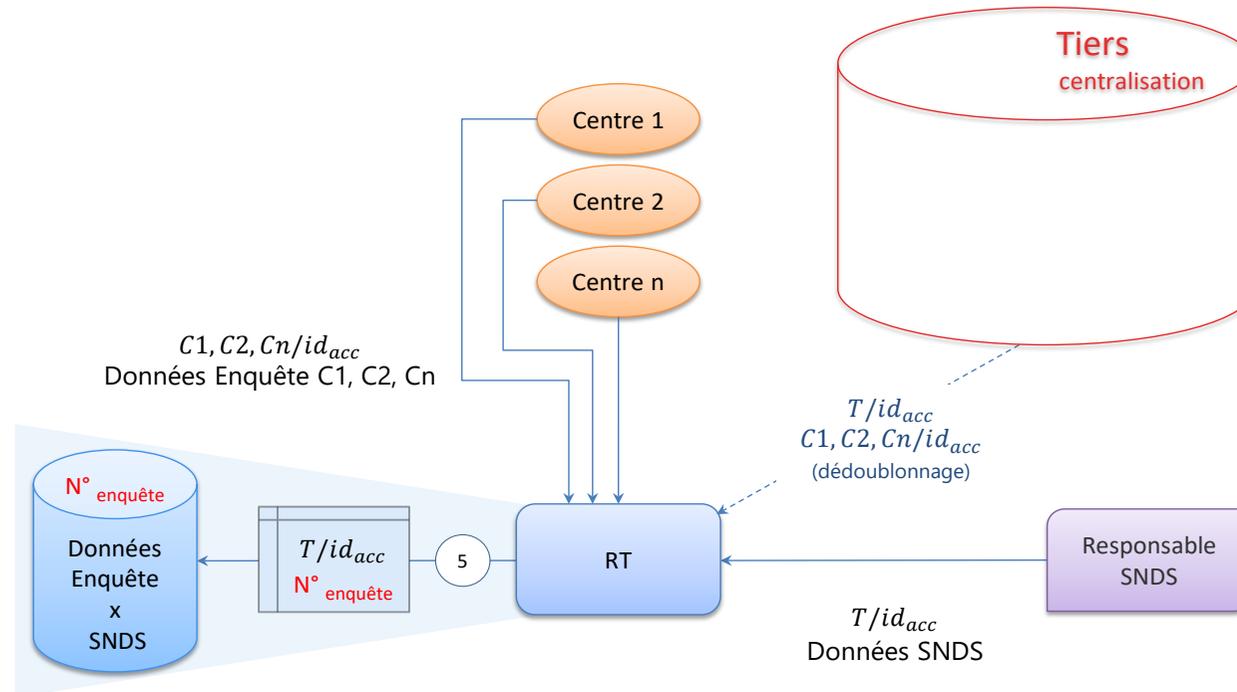
L'appariement doit être effectué sur une plateforme SNDS nationale ou dans un système (« bulle sécurisée ») conforme au [référentiel de sécurité du SNDS](#)

Le responsable de traitement reçoit les données du SNDS et les apparie avec les données d'enquête reçues des centres, à l'aide des identifiants d'accrochage

- Le cas échéant, il utilise la table de correspondance transmise par le tiers centralisateur, qui fait le lien entre les identifiants d'accrochage des centres et l'identifiant d'accrochage attribué par le tiers lors du dédoublement.

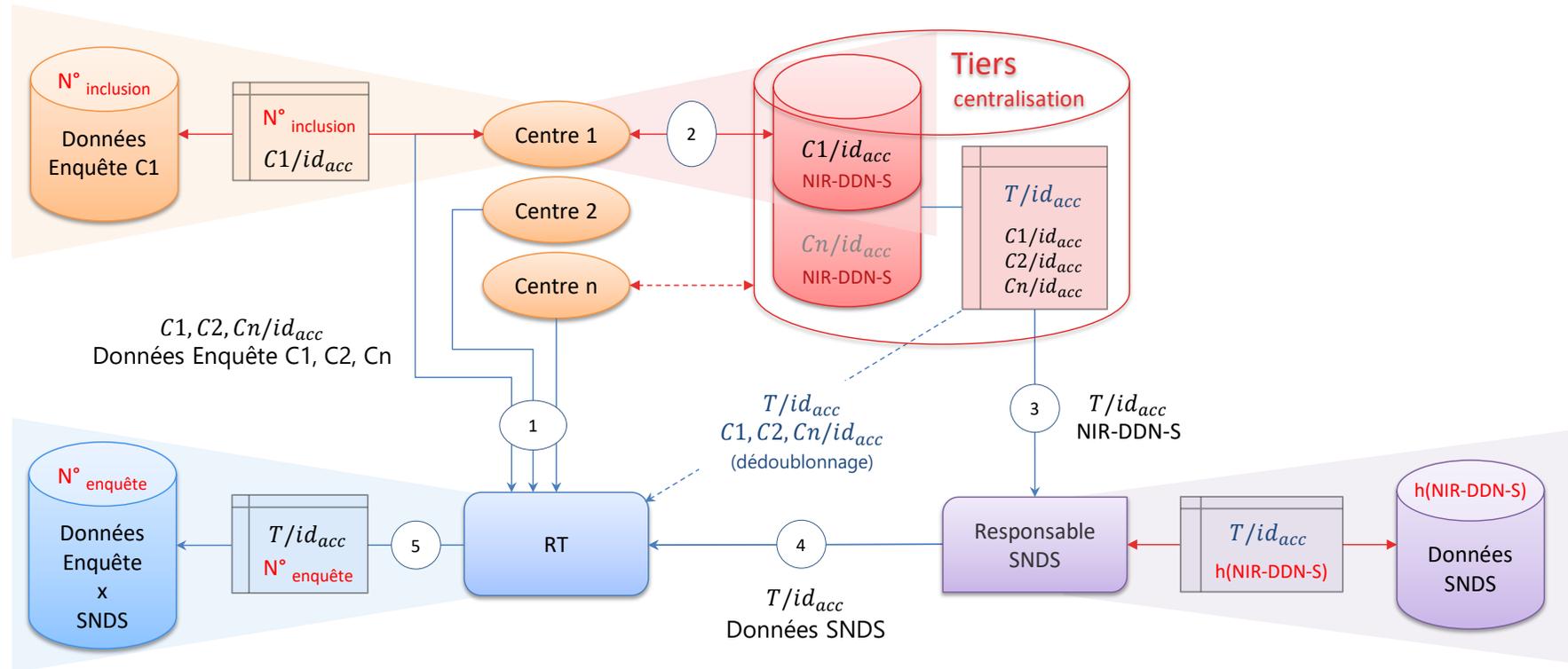
Après vérification de l'appariement, le responsable de traitement remplace les identifiants d'accrochage par un identifiant aléatoire propre à la base des données appariées : N° enquête

- Si le besoin est dûment justifié (par ex. transmission récurrente, audit des données, alerte des participants), la table de correspondance entre les identifiants d'accrochage et l'identifiant des données appariées peut être conservée par le responsable de traitement, de manière sécurisée.
- L'identifiant des données appariées peut être généré 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.



3. Synthèse de l'implémentation du circuit Multi-centres / NIR chez Tiers centralisateur

Vue fonctionnelle complète



Vue technique complète

