Fiche pratique

CIRCULATION DU NIR AUX FINS D'APPARIEMENT AVEC LE SNDS

CIRCUIT MULTI-CENTRES / eCRF SANS NIR

Produit en collaboration avec le CASD

MC - eCRF sans NIR – Version 1.0 – Juin 2024

CNIL.
COMMISSION NATIONALE
INFORMATIQUE & LIBERTÉS

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 CNIL[1] 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 vadémécum[2], comme aide-mémoire et guide de lecture.

La présente fiche concerne une étude multicentrique où les données des centres, sauf le NIR, sont gérées dans un « eCRF » opéré par le RT, et où le NIR reste stocké localement dans les centres.

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] *Vadémécum : 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

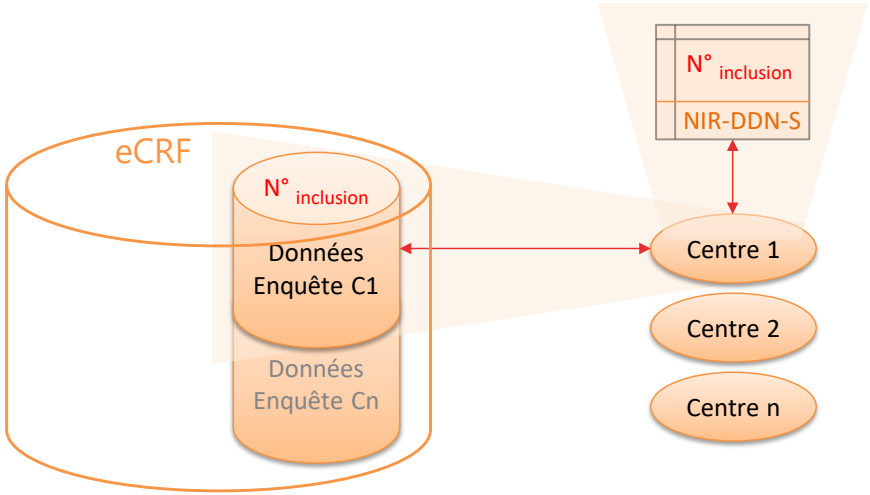
2. Implémentation détaillée du circuit Multi-centres / eCRF sans NIR

Saisie des données d'enquête

Chaque centre saisit ses données dans un eCRF mutualisé ne contenant pas le NIR

- Chaque centre n'accède qu'aux données qui lui sont propres.
- Le RT n'a pas d'accès direct aux données stockées dans le eCRF.
- Seul l'administrateur du eCRF peut accéder à l'ensemble des données.
- La sécurité est assurée par le eCRF selon les règles fixées par le RT.

- Afin de limiter les risques de réidentification, le eCRF est pseudonymisé et **le NIR des patients d'un centre reste stocké dans le centre, de manière cloisonnée par rapport au numéro d'inclusion** (par exemple, le NIR est chiffré avec une clé spécifique).
- En complément, **une authentification forte est recommandée pour tout accès au eCRF ainsi qu'à la table de correspondance locale** entre le numéro d'inclusion et le NIR, que ce soit en saisie de données ou pour une simple consultation.



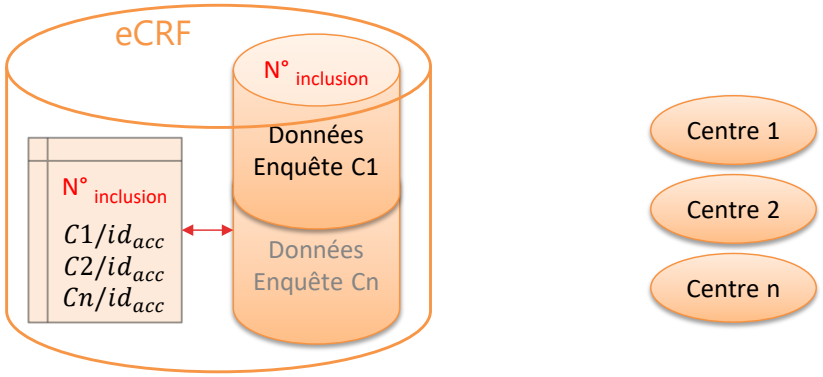
RT

Responsable SNDS

Etape 0 – Génération des identifiants d'accrochage

Le eCRF génère en interne, non visible des centres, une table de correspondance entre le numéro d'inclusion des participants et un identifiant d'accrochage aléatoire et non signifiant (C1/id_{acc} pour le premier centre, etc.)

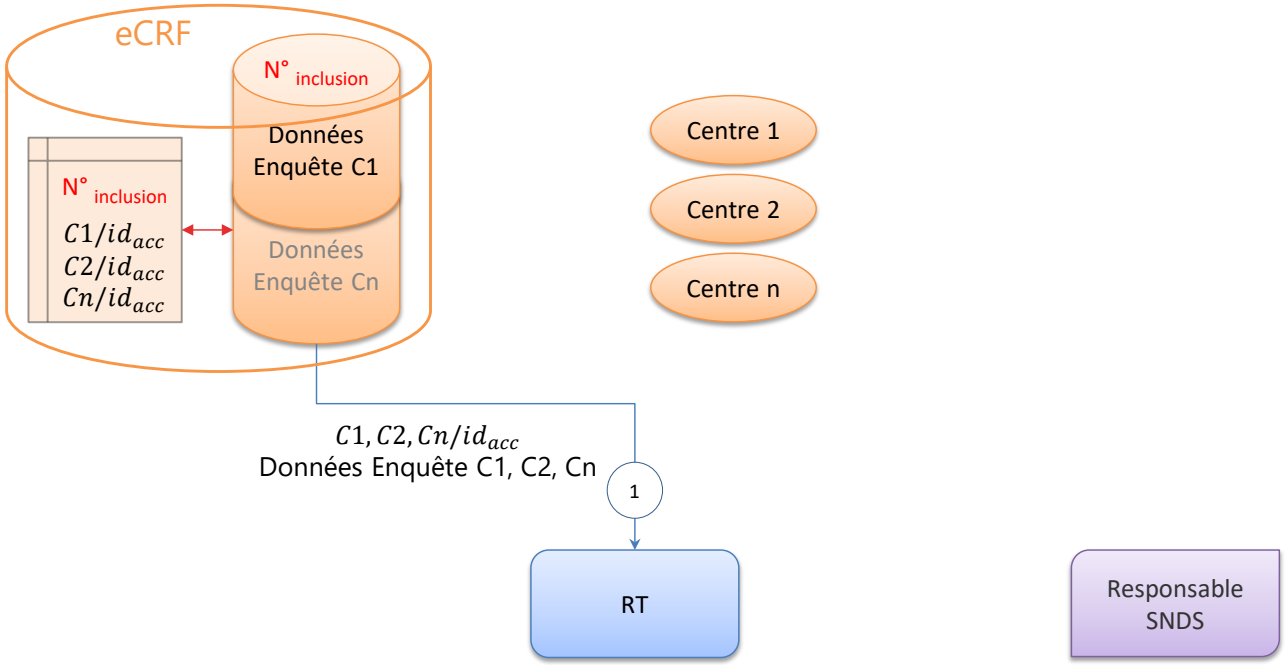
- L'utilisation d'identifiants techniques temporaires (« identifiants d'accrochage ») permet de dissocier le NIR et les données de santé lors des transferts entre acteurs.
- Par principe, **ces numéros sont non signifiants et différents du numéro d'inclusion de la personne dans l'étude**, afin de limiter les risques de réidentification croisée entre le numéro d'inclusion, le NIR et les données de santé (enquête et SNDS).
- De même, **les centres n'ont pas d'accès direct à la table de correspondance interne au eCRF** qui fait le lien entre les numéros d'inclusion et les identifiants d'accrochage.
- 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 dans le eCRF, 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.
- Si l'identifiant d'accrochage est généré au fil des inclusions, il doit rester masqué dans l'interface eCRF des centres.



Etape 1 – Envoi des données d'enquête au RT

Le eCRF envoie au responsable de traitement les données d'enquête de chaque centre, 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 numéro d'inclusion n'est pas transmis au responsable de traitement**.
- **À noter :** 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 le eCRF** : pour extraire les données strictement nécessaires (identifiants d'accrochage et données d'enquête), il peut demander l'intervention manuelle d'un administrateur habilité ou déclencher une fonction interne au eCRF.
- Dans le cas d'une fonction interne au eCRF, celle-ci pourra également envoyer aux centres les données qui leur sont nécessaires pour le circuit d'appariement (cf. étape 2).

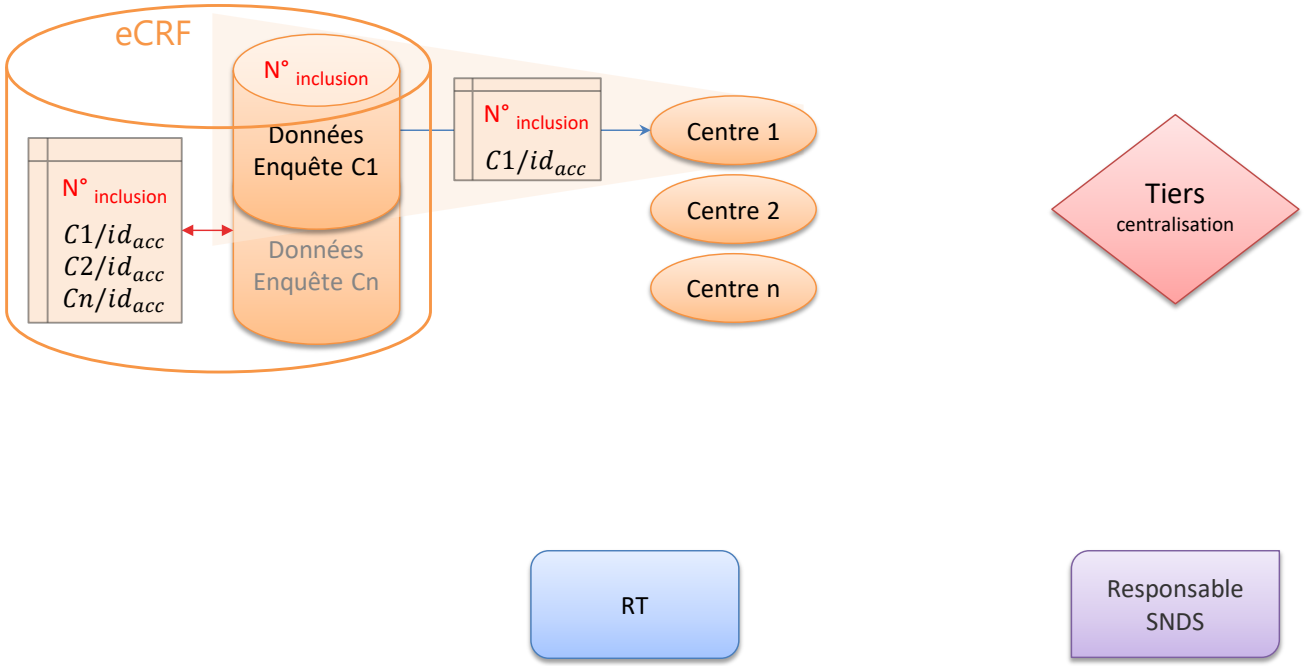


Etape 2 – Envoi des NIR au tiers centralisateur

Chaque centre reçoit la liste des numéros d'inclusion de ses participants, accompagnés des identifiants d'accrochage correspondants

- Par principe, **les centres n'ont pas d'accès complet à la table de correspondance globale du eCRF** entre les numéros d'inclusion et les identifiants d'accrochage. En effet, un centre ne peut être destinataire que des données qui lui sont strictement nécessaires pour le circuit d'appariement, à savoir les numéros d'inclusion de ses seuls patients et les identifiants d'accrochage associés.

- L'extraction de ces données va donc être réalisée pour chaque centre, soit manuellement par un administrateur habilité du eCRF, à la demande du RT, soit par une fonction interne au eCRF qui sera déclenchée par le RT (cf. étape 1), soit par une fonction interne déclenchée par chaque centre dans son interface eCRF.

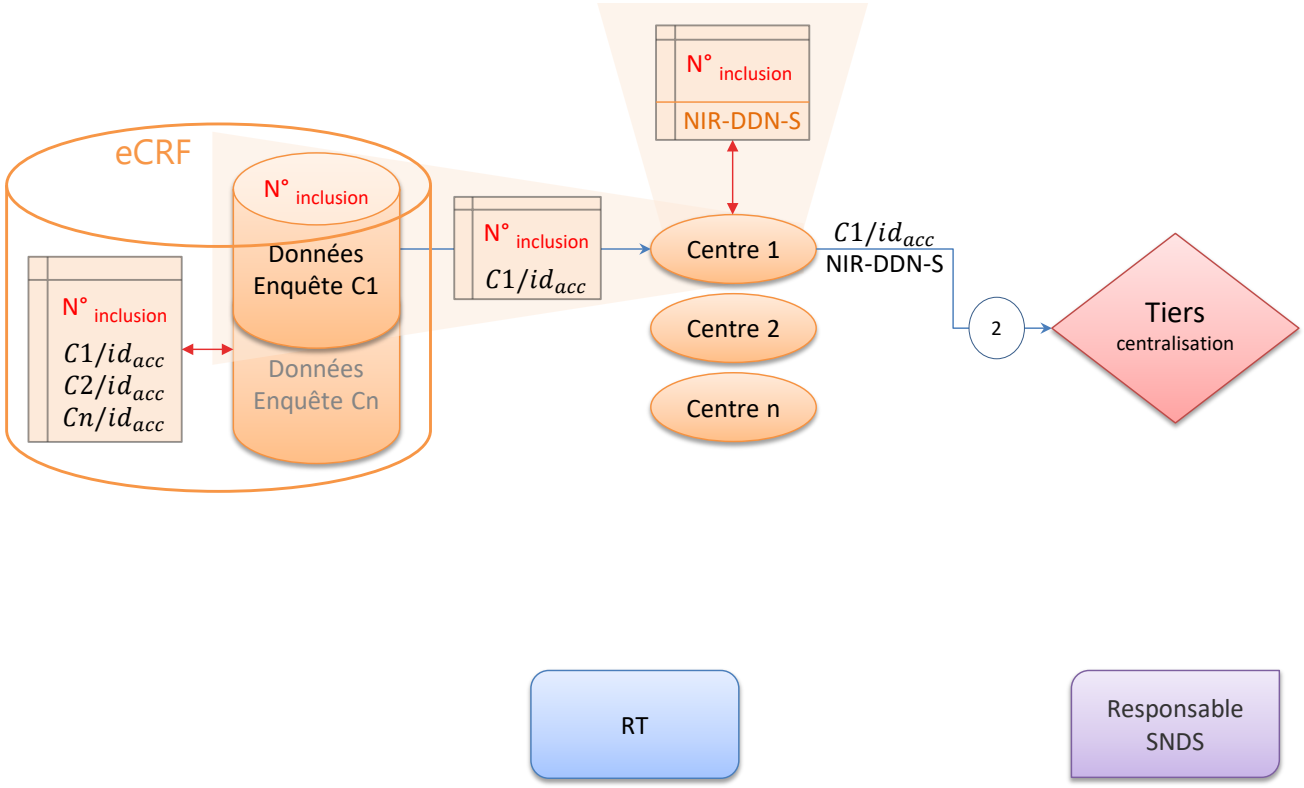


Etape 2 – Envoi des NIR au tiers centralisateur (suite)

Chaque centre transmet au tiers centralisateur les identifiants d'accrochage accompagnés des [NIR - Date de Naissance - Sexe] de ses participants

- À partir des numéros d'inclusion reçus et à l'aide de sa table de correspondance locale, chaque centre va générer un fichier associant les identifiants d'accrochage reçus avec les [NIR - Date de Naissance - Sexe] correspondants. Ce fichier va être transmis au tiers centralisateur, comme seules données strictement nécessaires au circuit d'appariement.

- En effet, par principe, **le numéro d'inclusion 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 dans les centres, et aucune autre donnée personnelle liée à l'enquête ne lui est transmise.
- Enfin, **le centre ne conserve pas le fichier contenant les NIR**, qui doit être détruit juste après l'envoi au tiers, **de même que celui qui contenait les numéros d'inclusion**.

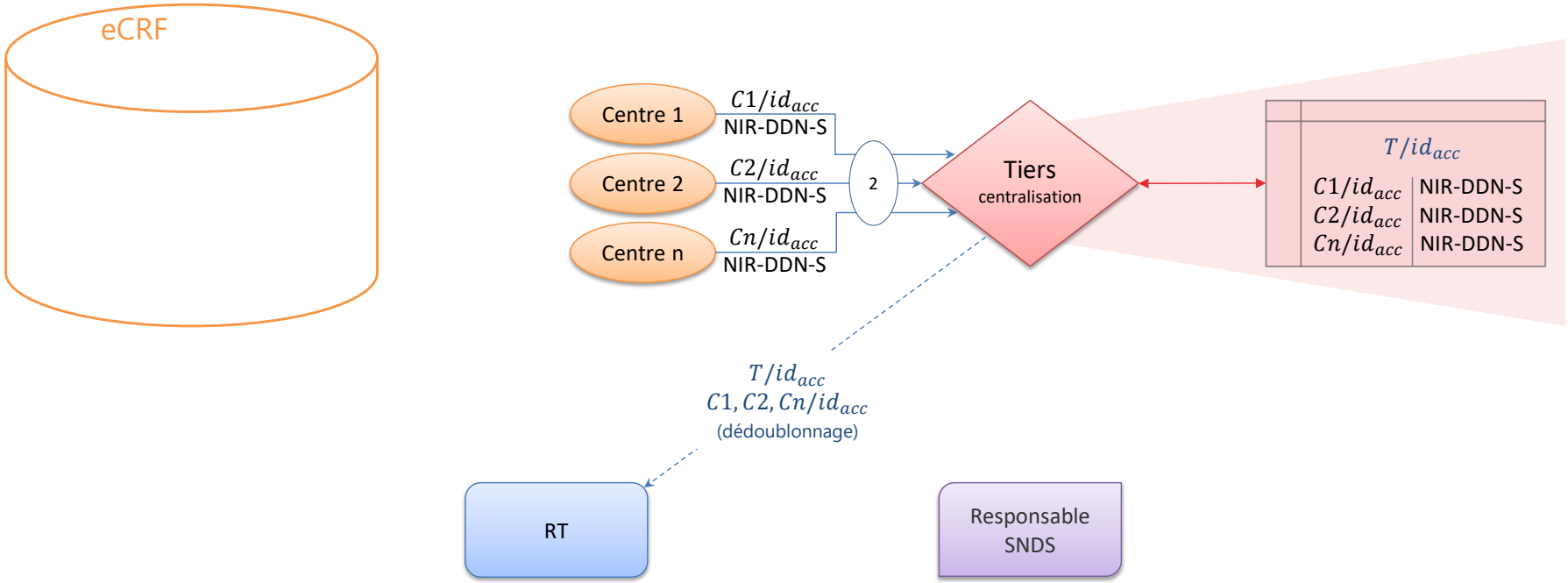


Etape 2bis – Centralisation, dédoublonnage et envoi au RT

Le tiers centralisateur fusionne les tables reçues de chaque centre

Si nécessaire, il procède 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}

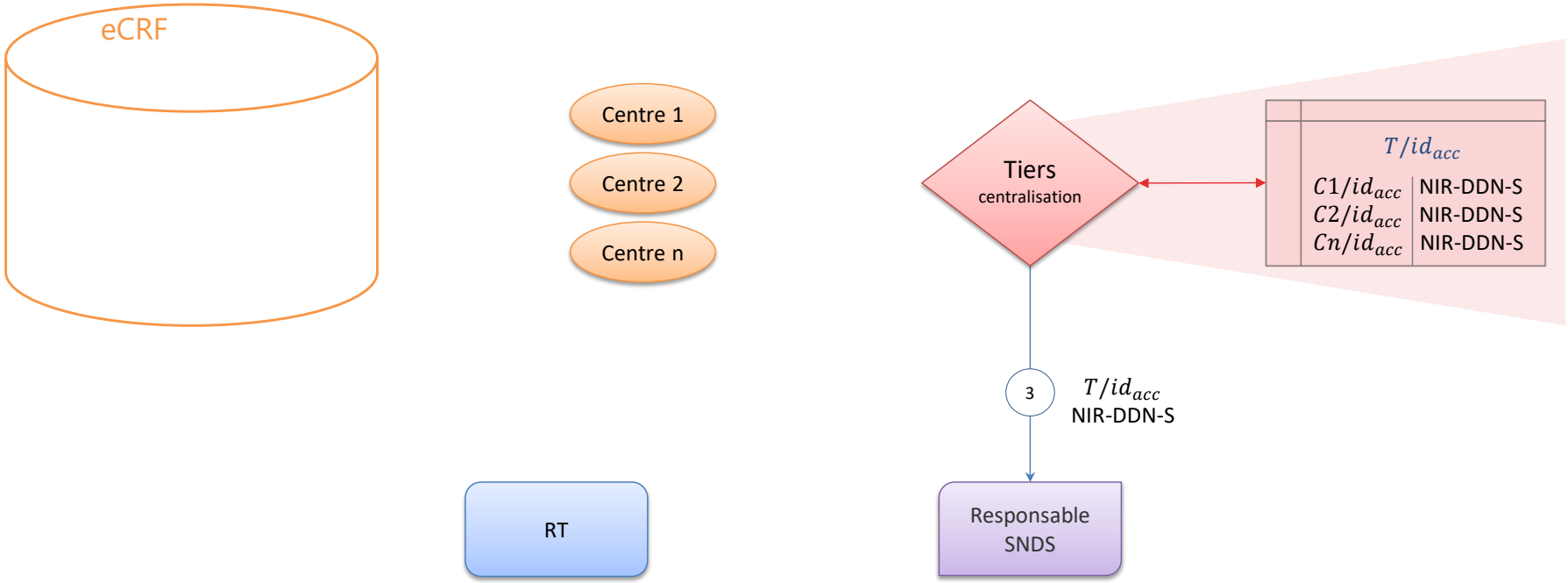


Etape 3 – Envoi des NIR au responsable SNDS

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.

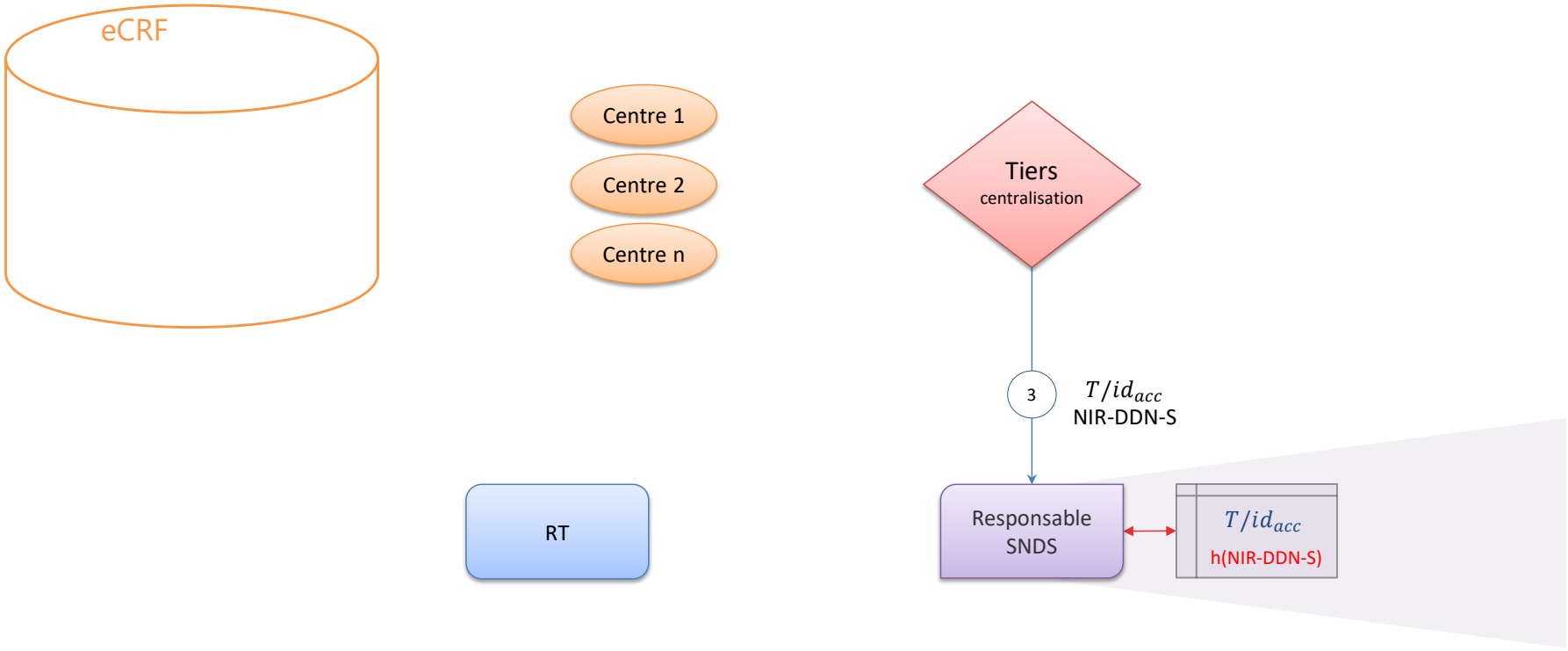


Etape 3bis – Hachage des NIR en entrée

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(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(NIR-DDN-S) dès réception des données et il n'est pas conservé, afin de limiter les risques de réidentification des données du SNDS.

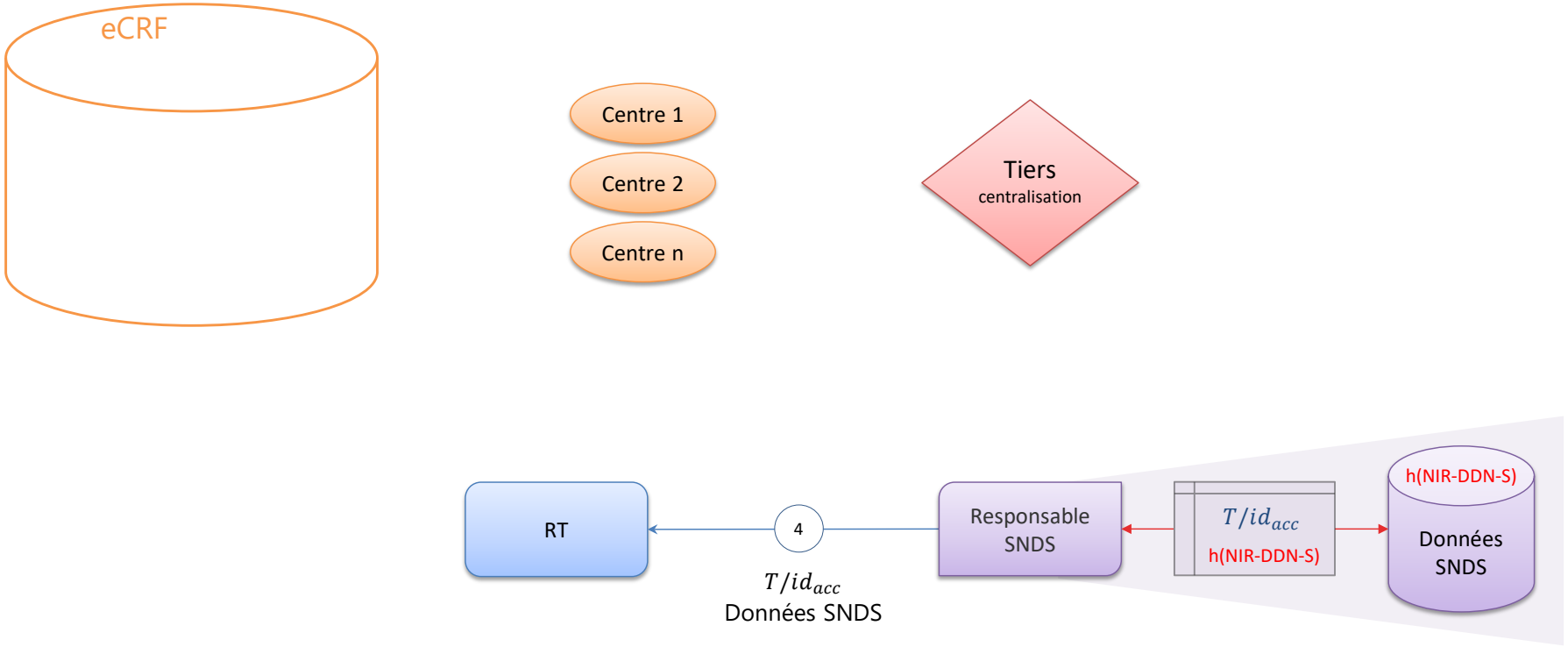


Etape 4 – Extraction et envoi des données SNDS au RT

Le responsable de la base SNDS extrait les données du SNDS correspondant aux h(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(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

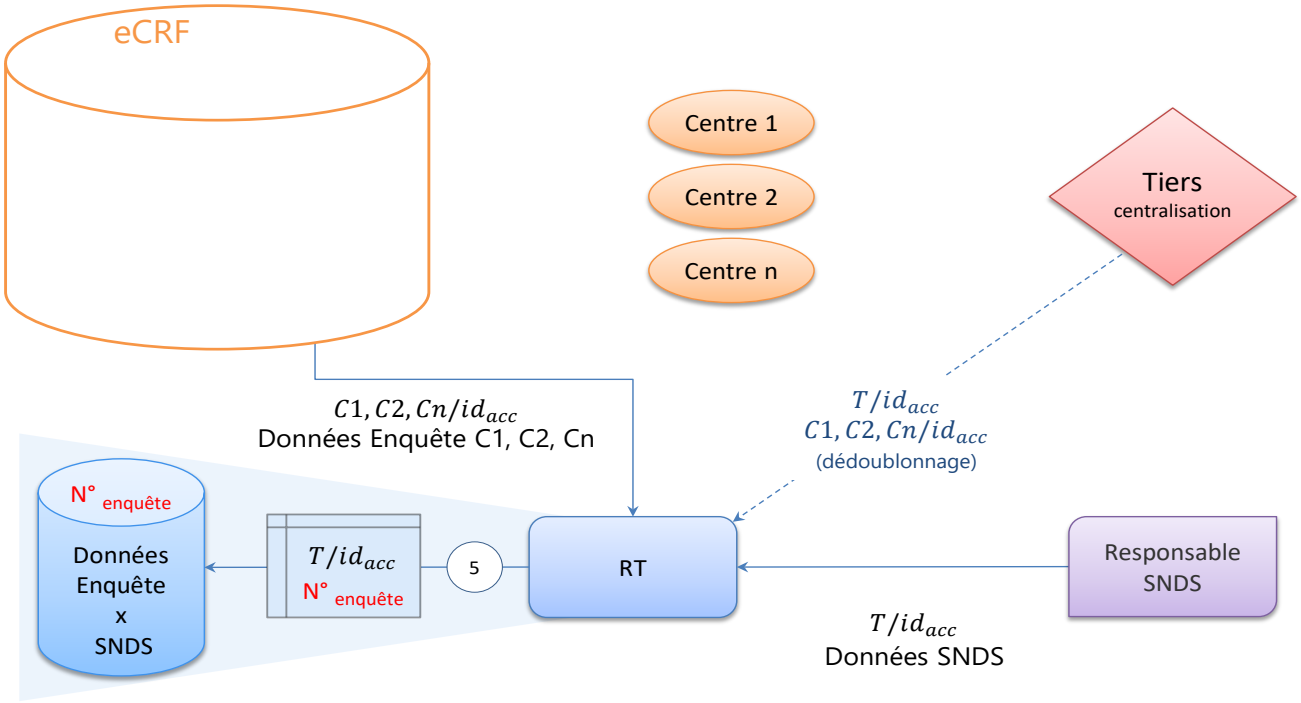
L'appariement doit être effectué sur une plateforme SNDS nationale ou dans un système (« bulle sécurisée ») conforme au référentiel de sécurité du SNDS

Le responsable de traitement reçoit les données du SNDS et les apparie avec les données d'enquête transmises par le eCRF, 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 dédoublonnage.

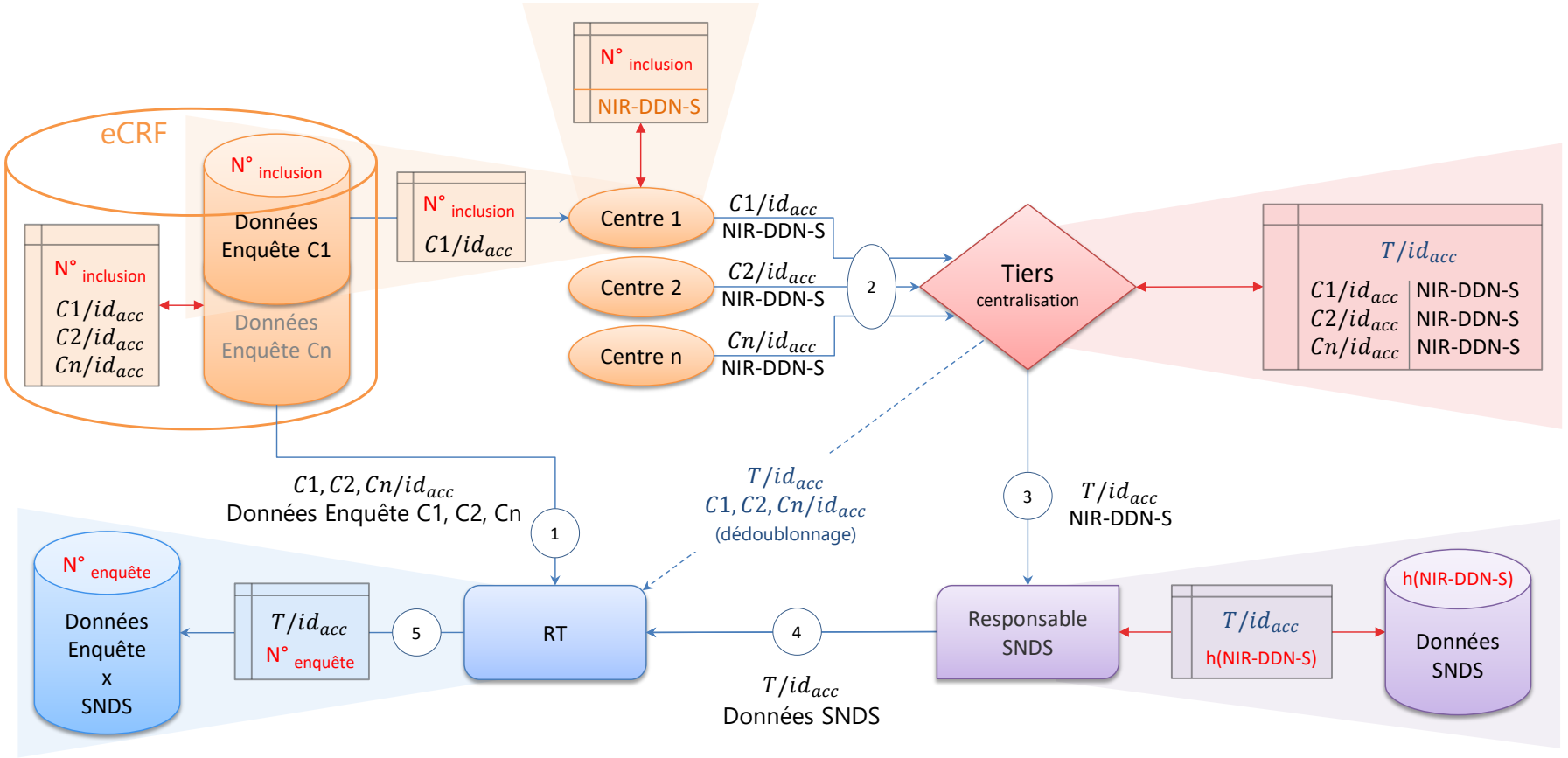
Après vérification de l'appariement, le responsable de traitement remplace les identifiants d'accrochage par un identifiant aléatoire propre à la base des données appariées : N° enquête

- 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 / eCRF sans NIR

Vue fonctionnelle complète



Vue technique complète