Fiche pratique

CIRCULATION DU NIR AUX FINS D'APPARIEMENT AVEC LE SNDS

CIRCUIT MULTI-CENTRES / eCRF AVEC NIR

Produit en collaboration avec le CASD

MC - eCRF avec 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, y compris le NIR, sont gérées dans un « eCRF » opéré par le RT.

Un tiers de confiance peut être sollicité pour mettre au format et opérer la transmission du NIR à la CNAM.

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

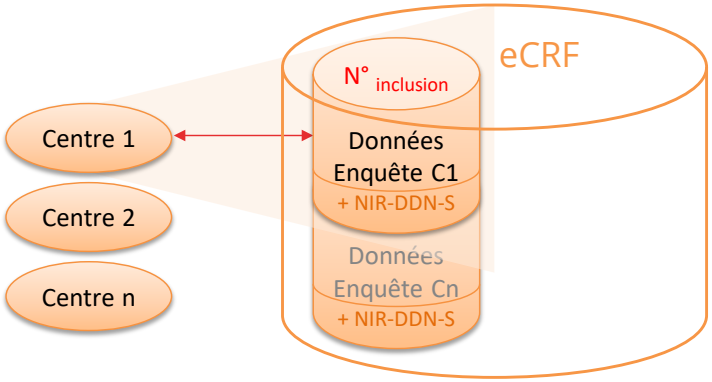
2. Implémentation détaillée du circuit Multi-centres / eCRF avec NIR

Saisie des données d'enquête

Chaque centre saisit ses données dans un eCRF mutualisé contenant 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.

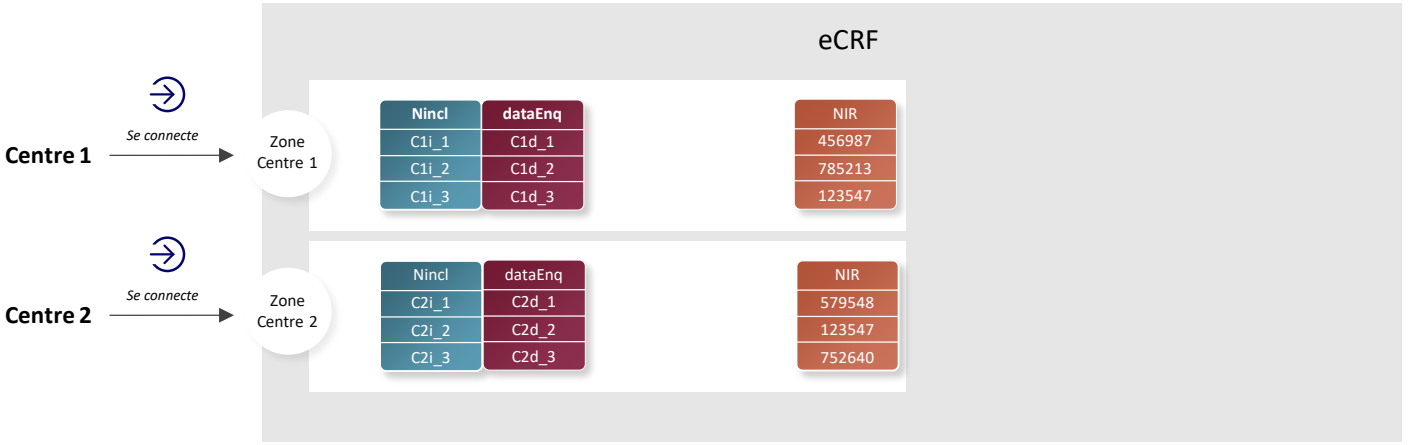
- Afin de limiter les risques de réidentification, **le NIR est cloisonné par rapport aux données d'enquête et au numéro d'inclusion** (par exemple, dans une table séparée et chiffrée avec une clé spécifique).
- En raison des risques liés aux accès à distance via Internet et à la présence dans le eCRF des données identifiantes [NIR - Date de Naissance - Sexe], **une authentification forte est exigée à chaque connexion au eCRF**.



RT

Responsable SNDS

Saisie des données d'enquête

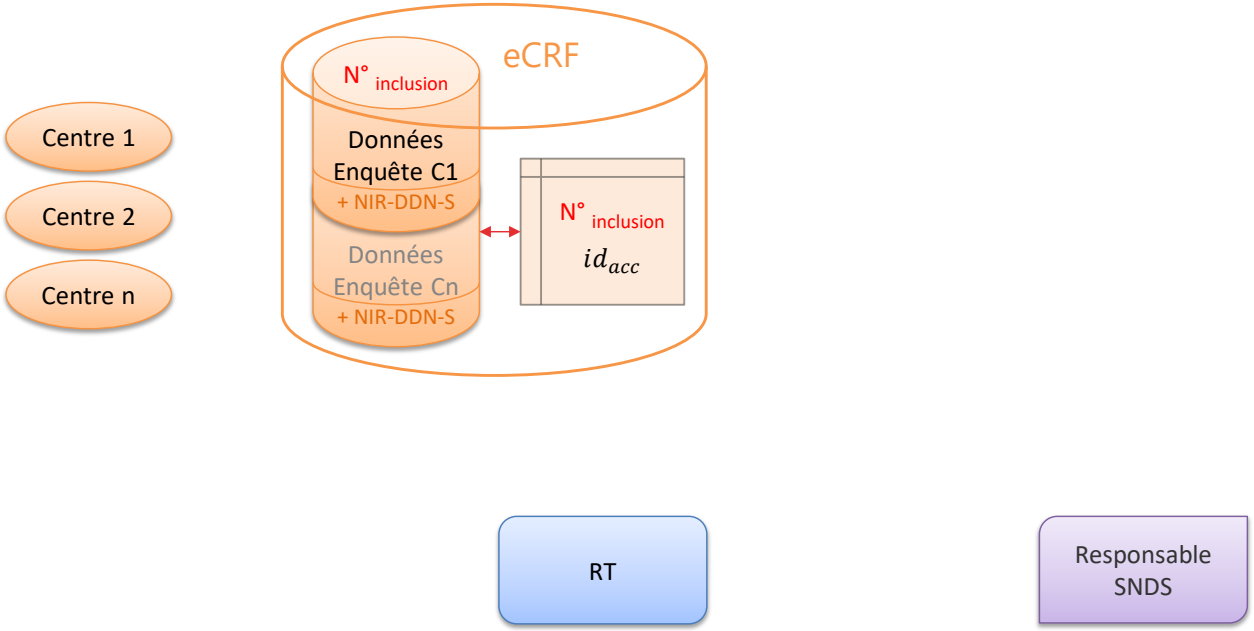


La colonne **Nincl** représente le numéro d'inclusion
La colonne **dataEnq** représente les données collectée pour l'enquête ou le registre ou la cohorte…
La colonne **NIR** représente le triplet NIR -Date de Naissance –Sexe

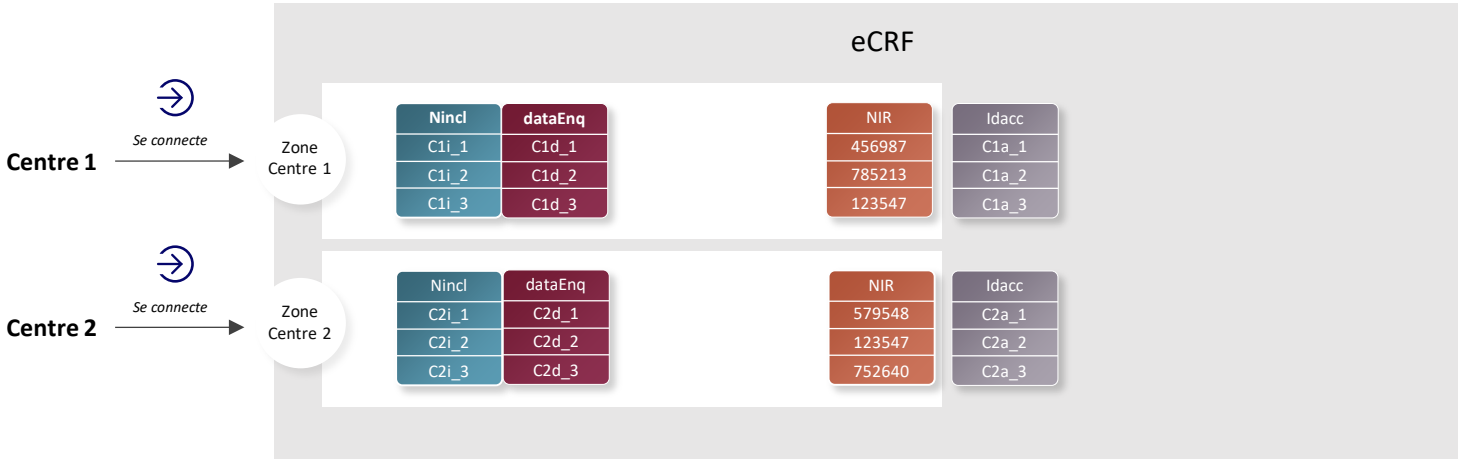
Etape 0 – Génération des identifiants d'accrochage

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 (id_{acc})

- 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 accès 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 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 0 – Génération des identifiants d'accrochage

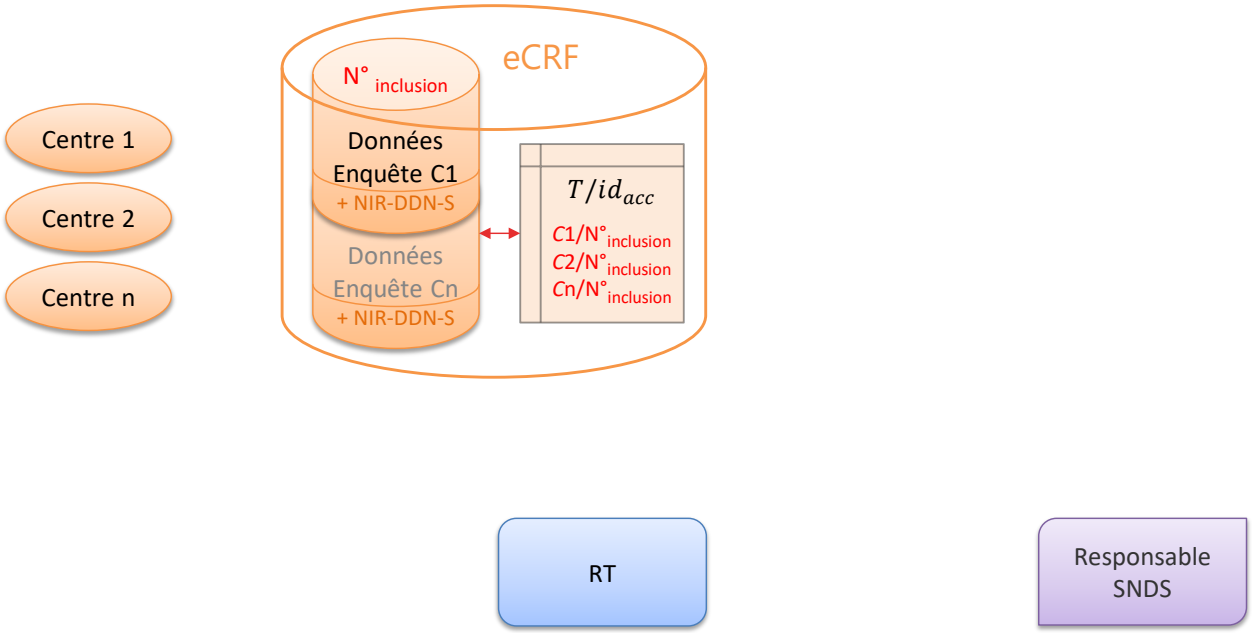


La colonne **Nincl** représente le numéro d'inclusion
La colonne **dataEnq** représente les données collectée pour l'enquête ou le registre ou la cohorte…
La colonne **NIR** représente le triplet NIR -Date de Naissance –Sexe
La colonne **Idacc** représente le numéro d'accrochage

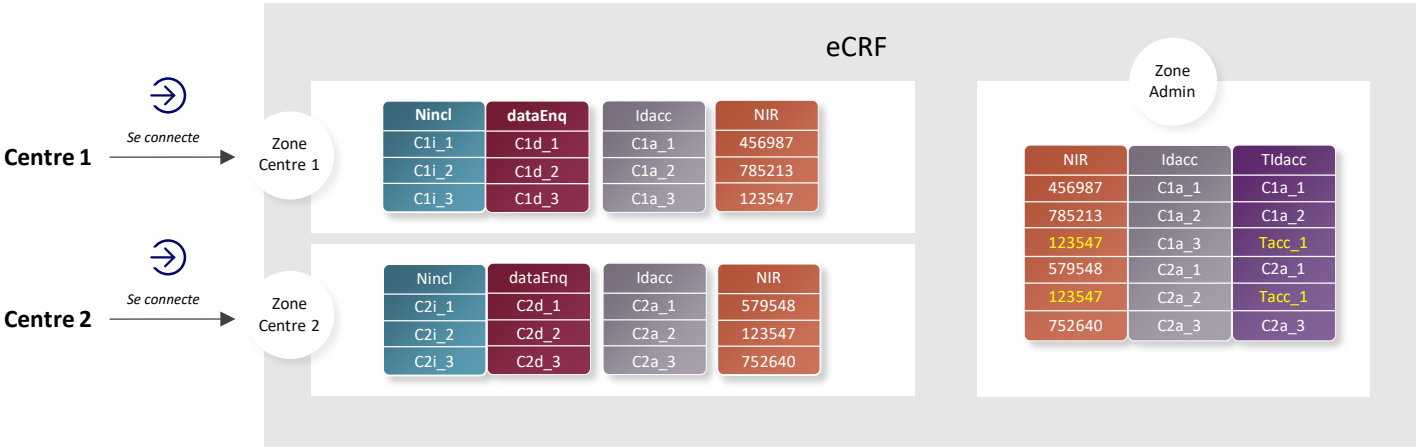
Etape 0bis – Dédoublonnage

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 dès lors plusieurs numéros d'accrochage id_{acc} .
- S'il est nécessaire de chaîner les données issues de plusieurs centres, le eCRF identifie les lignes correspondant à un même [NIR - Date de Naissance - Sexe] et ne conserve qu'un identifiant d'accrochage id_{acc} unique pour les différents numéros d'inclusion d'un même participant.
- S'il n'y a pas besoin de dédoublonnage, il utilise les id_{acc} attribués pour chaque numéro d'inclusion (cf. étape 0).



Etape 0bis – Dédoublonnage

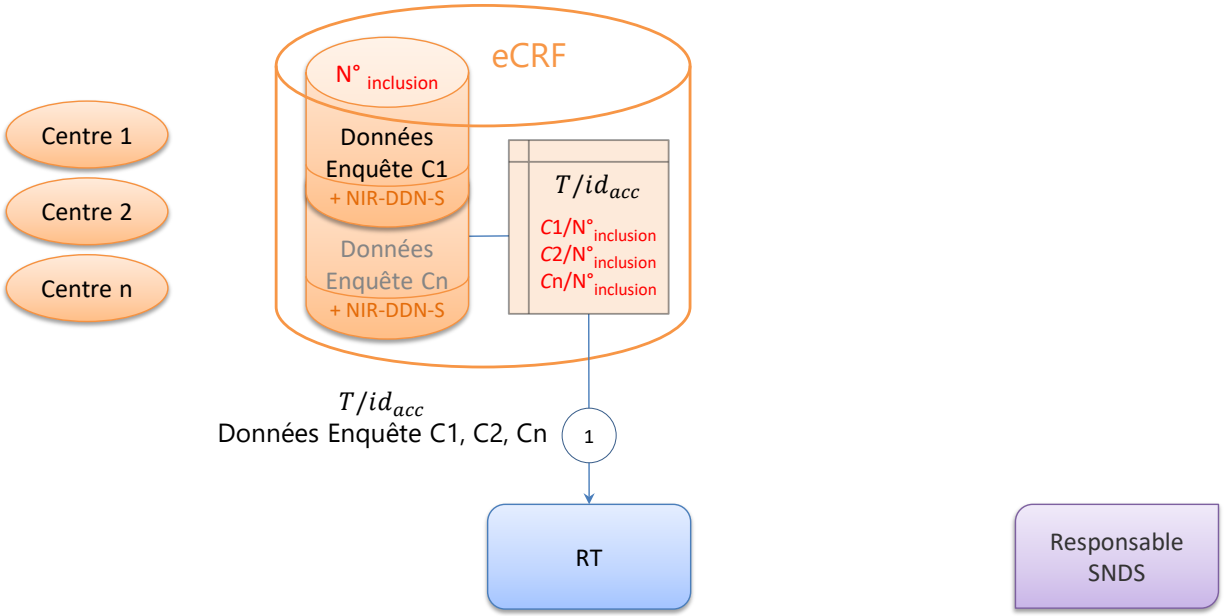


La colonne **Nincl** représente le numéro d'inclusion
La colonne **dataEnq** représente les données collectée pour l'enquête ou le registre ou la cohorte…
La colonne **Idacc** représente le numéro d'accrochage
La colonne **NIR** représente le triplet NIR -Date de Naissance –Sexe
La colonne **Tidacc** représente le dédoublonnage de la colonne **Idacc**

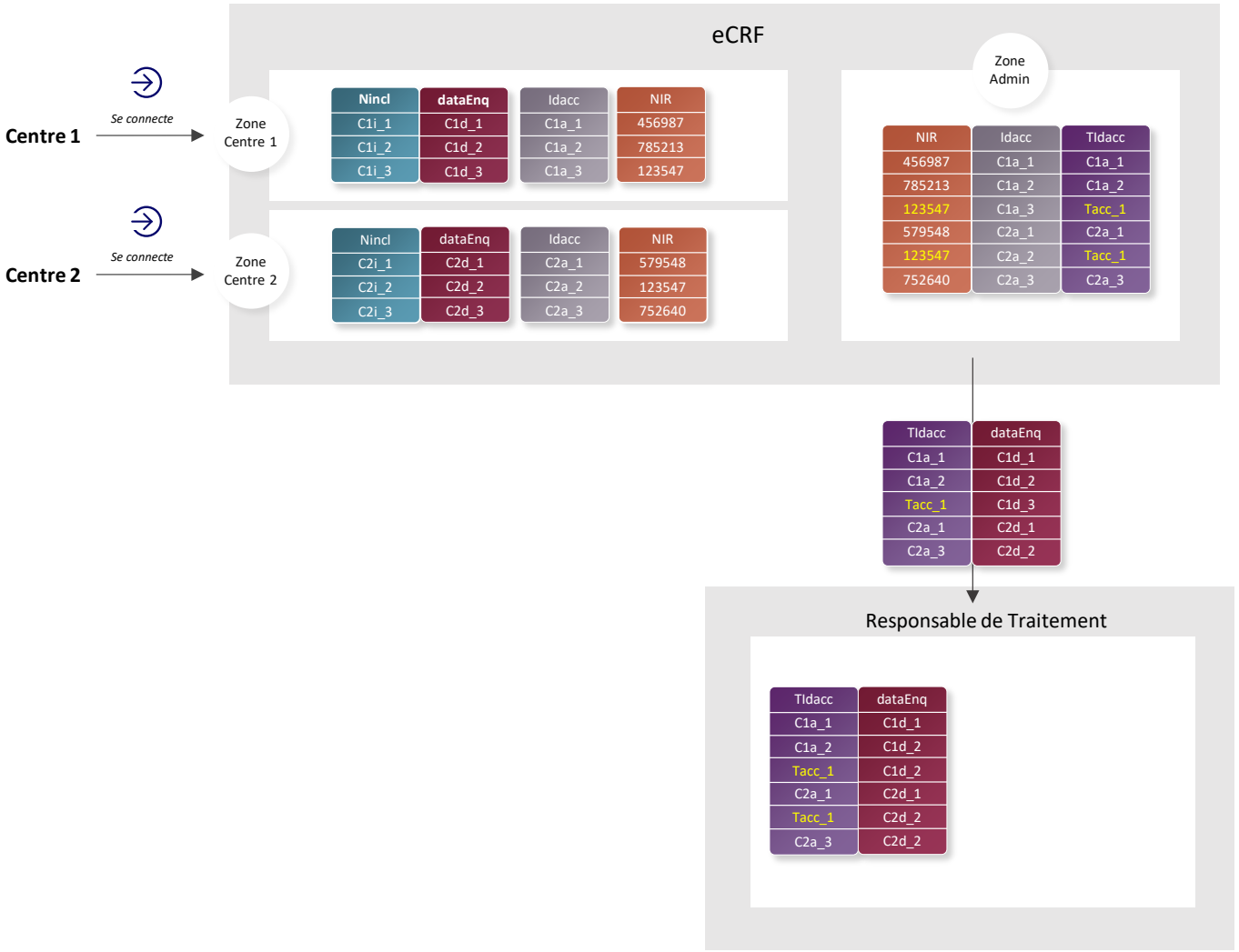
Etape 1 – Envoi des données d'enquête au RT

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 NIR et le numéro d'inclusion ne sont 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 au tiers de formatage les données qui lui sont nécessaires pour le circuit d'appariement (cf. étape 2).



Etape 1 – Envoi des données d'enquête au RT

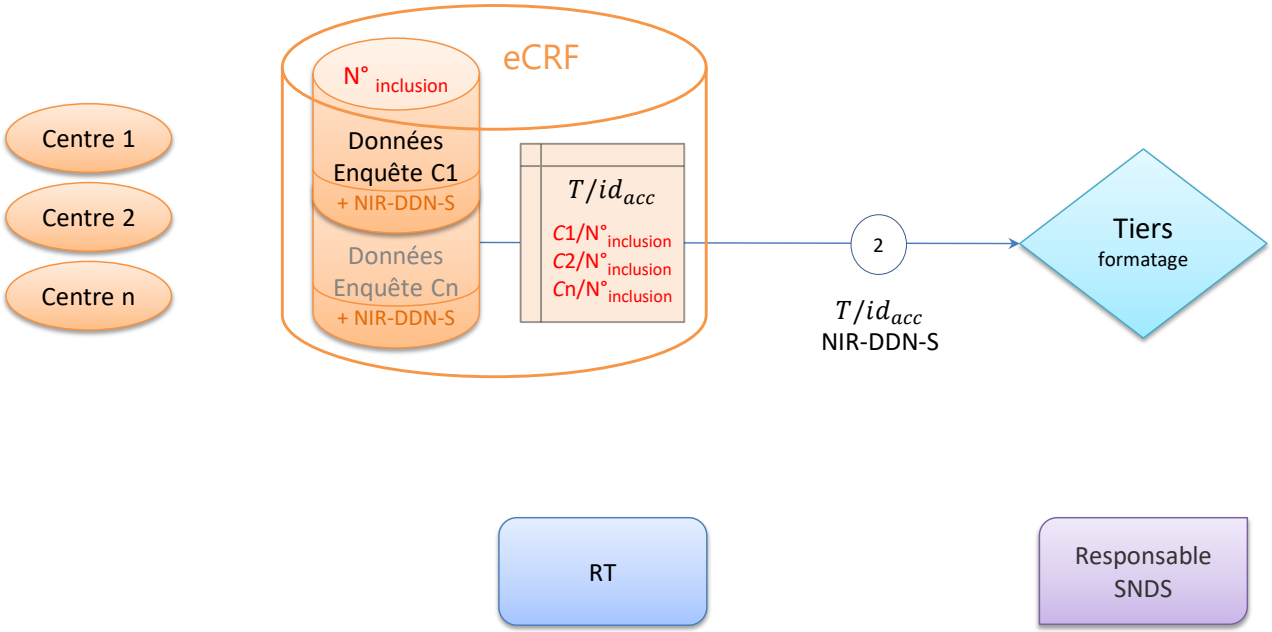


Etape 2 – Envoi des NIR au tiers de formatage

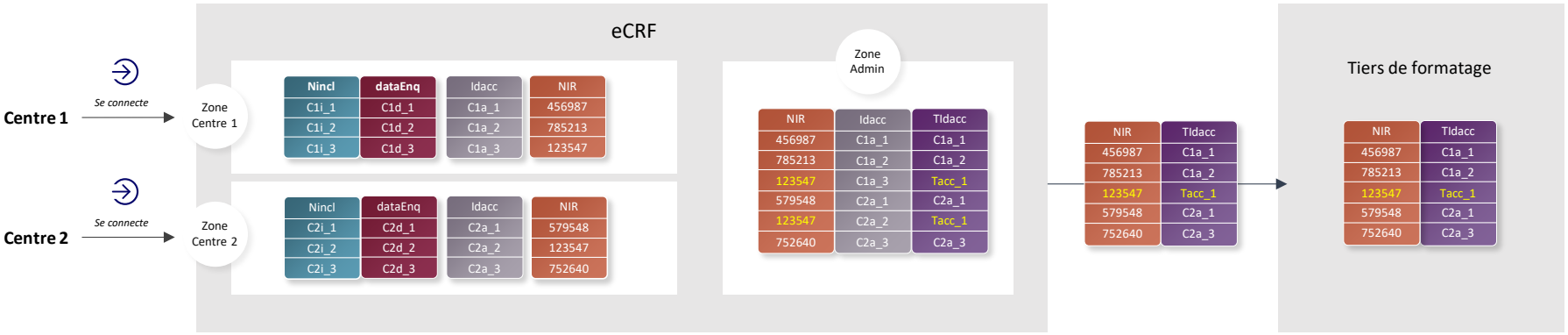
Le eCRF transmet au tiers de formatage les identifiants d'accrochage accompagnés des [NIR - Date de Naissance - Sexe] des participants de l'ensemble des centres

- Le RT va demander l'extraction des données à transmettre au tiers de formatage, qui sera effectuée soit manuellement, par un administrateur habilité, soit par déclenchement d'une fonction interne au eCRF (cf. étape 1). Un fichier associant les identifiants d'accrochage avec les [NIR - Date de Naissance - Sexe] correspondants sera alors généré et transmis au tiers, 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 de formatage n'a pas d'accès direct** aux données stockées dans le eCRF, et aucune autre donnée personnelle liée à l'enquête ne lui est transmise.
- Enfin, **le eCRF ne conserve pas le fichier contenant les NIR**, qui doit être détruit juste après l'envoi au tiers.



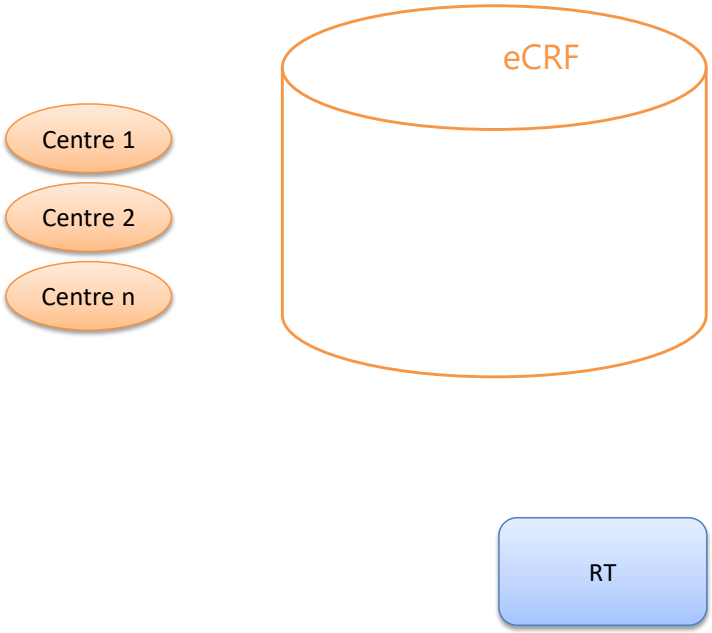
Etape 2 – Envoi des NIR au tiers de formatage

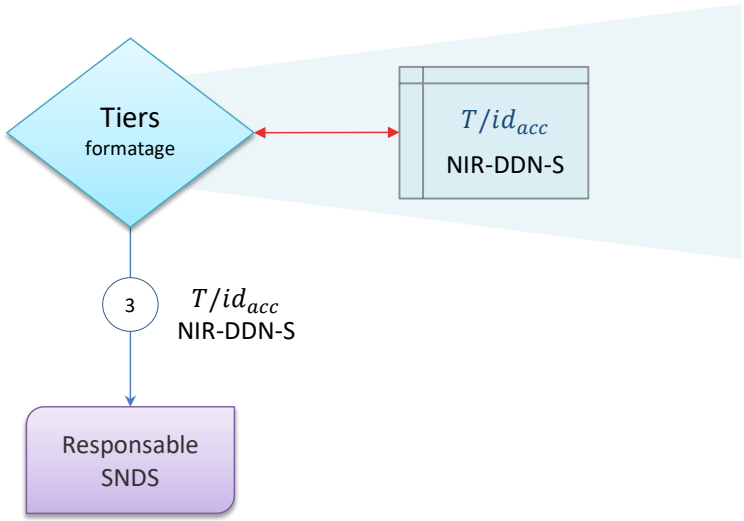


Etape 3 – Envoi des NIR au responsable SNDS

Le tiers de formatage envoie au responsable de la base SNDS les 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.
- Le tiers a ici pour rôle de formater les données reçues du eCRF et d'opérer le téléservice « SAFE ».
- Si le besoin est dûment justifié (par ex. transmission récurrente, audit des données, alerte des participants), le tiers peut conserver le fichier transmis, de manière sécurisée.





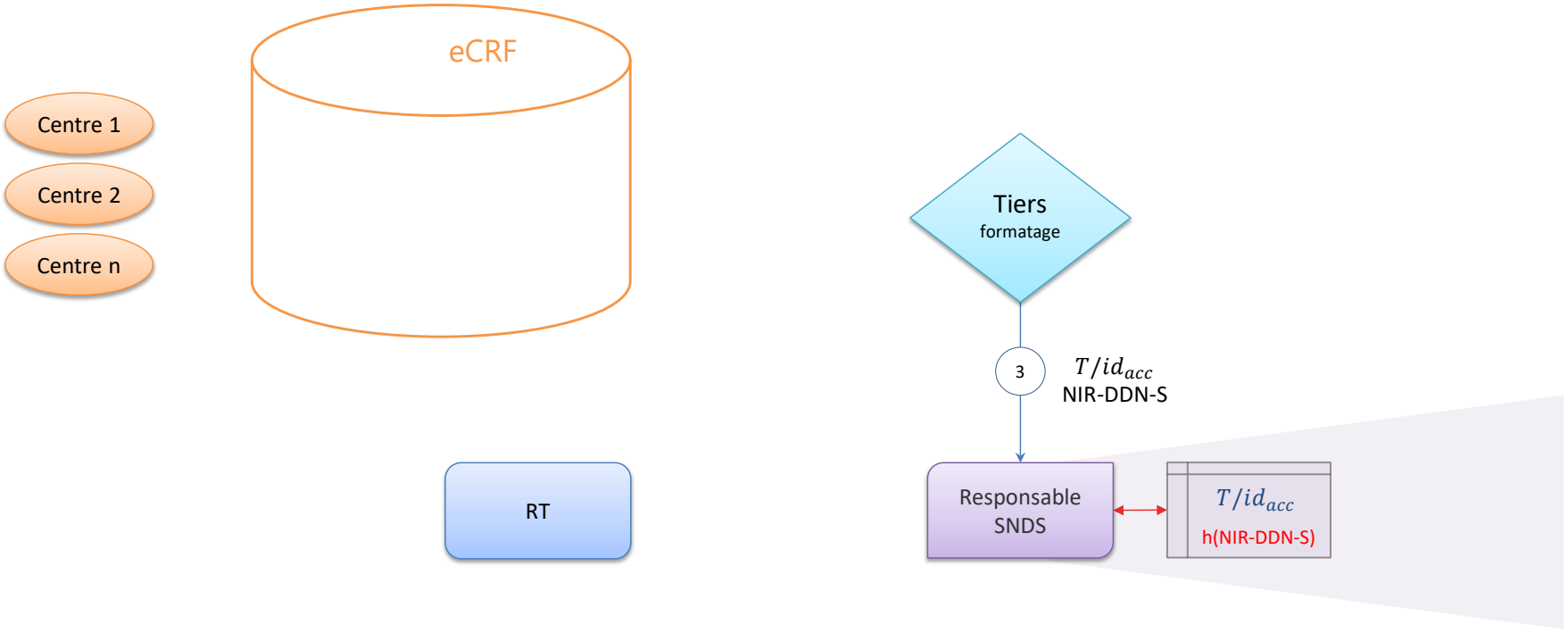
Etape 3 – Envoi des NIR au responsable SNDS

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.



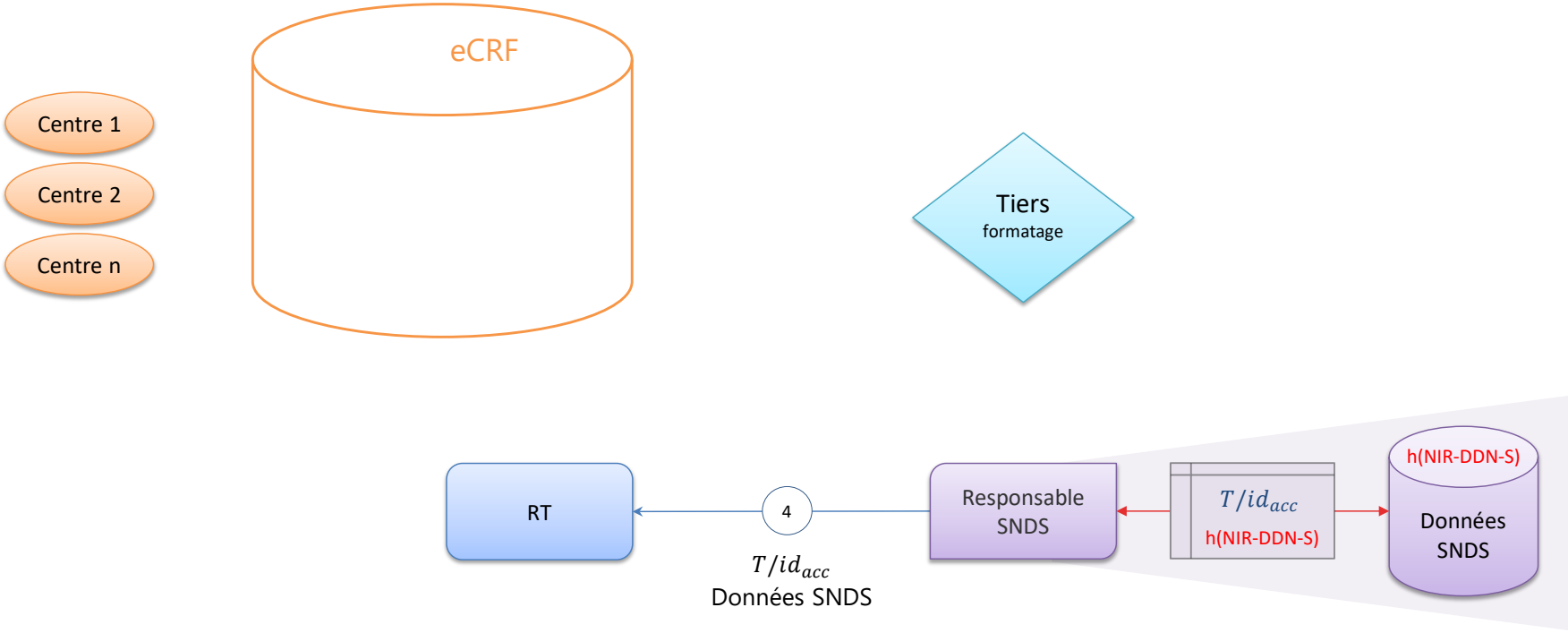
Etape 3bis – Hachage des NIR en entrée

Etape 4 – Extraction et envoi des données SNDS au RT

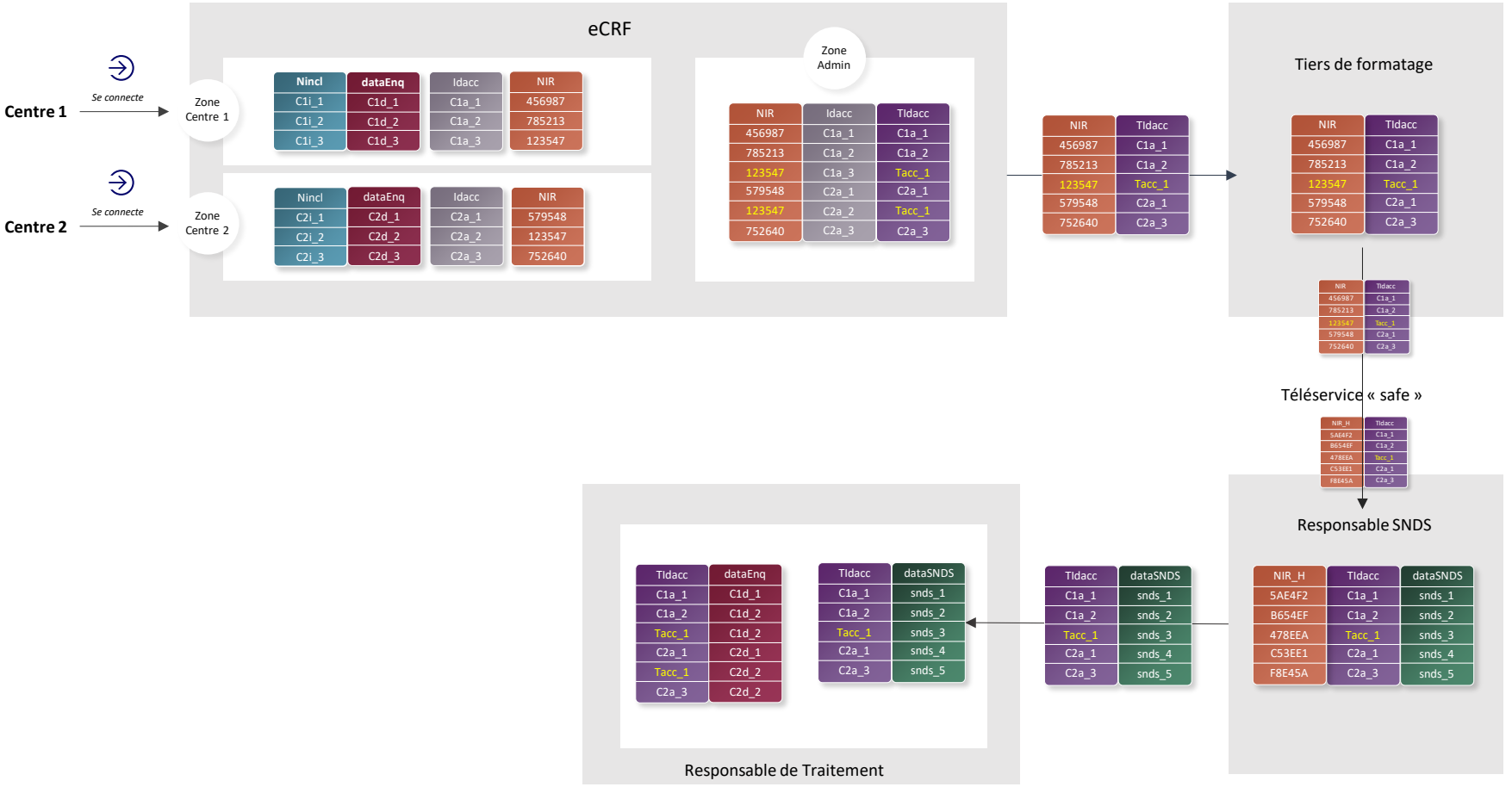
Le responsable de la base SNDS extrait les données du SNDS correspondant aux h(NIR-DDN-S) des participants

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

- 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 4 – Extraction et envoi des données SNDS au RT



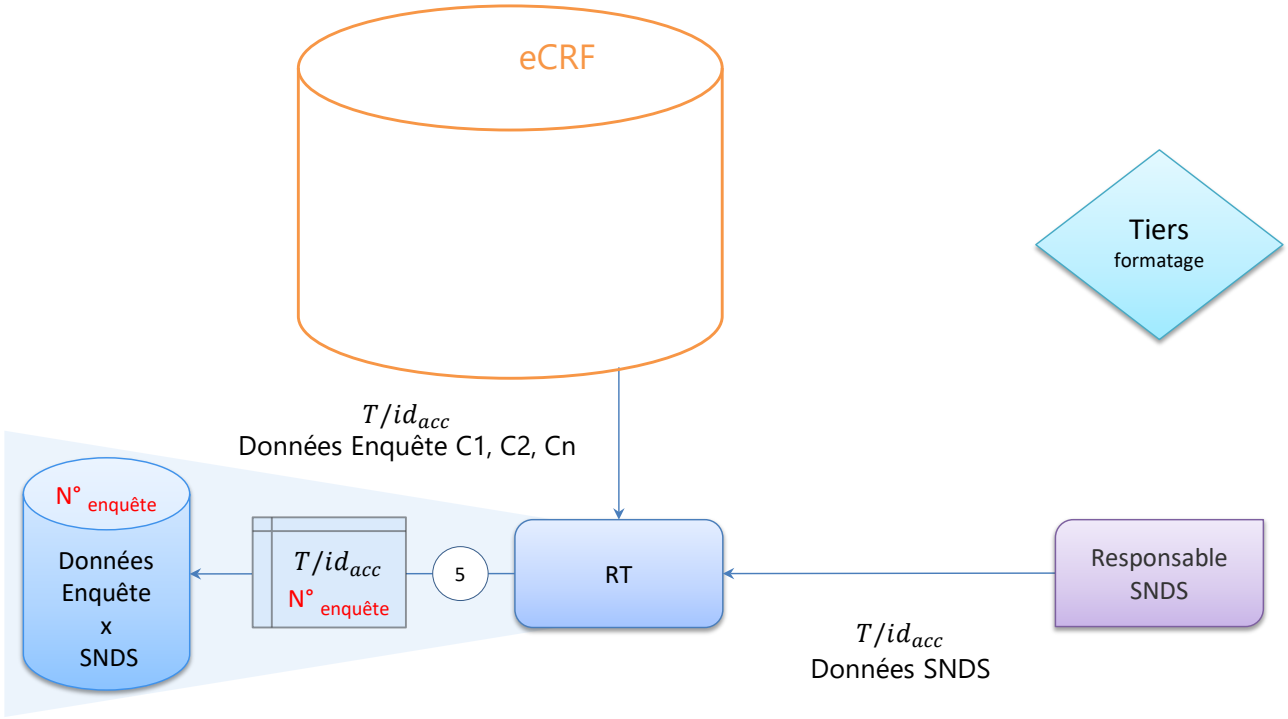
Etape 5 – Appariement des données d'enquête avec les données du SNDS, et génération d'un nouvel identifiant pour la base appariée

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

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

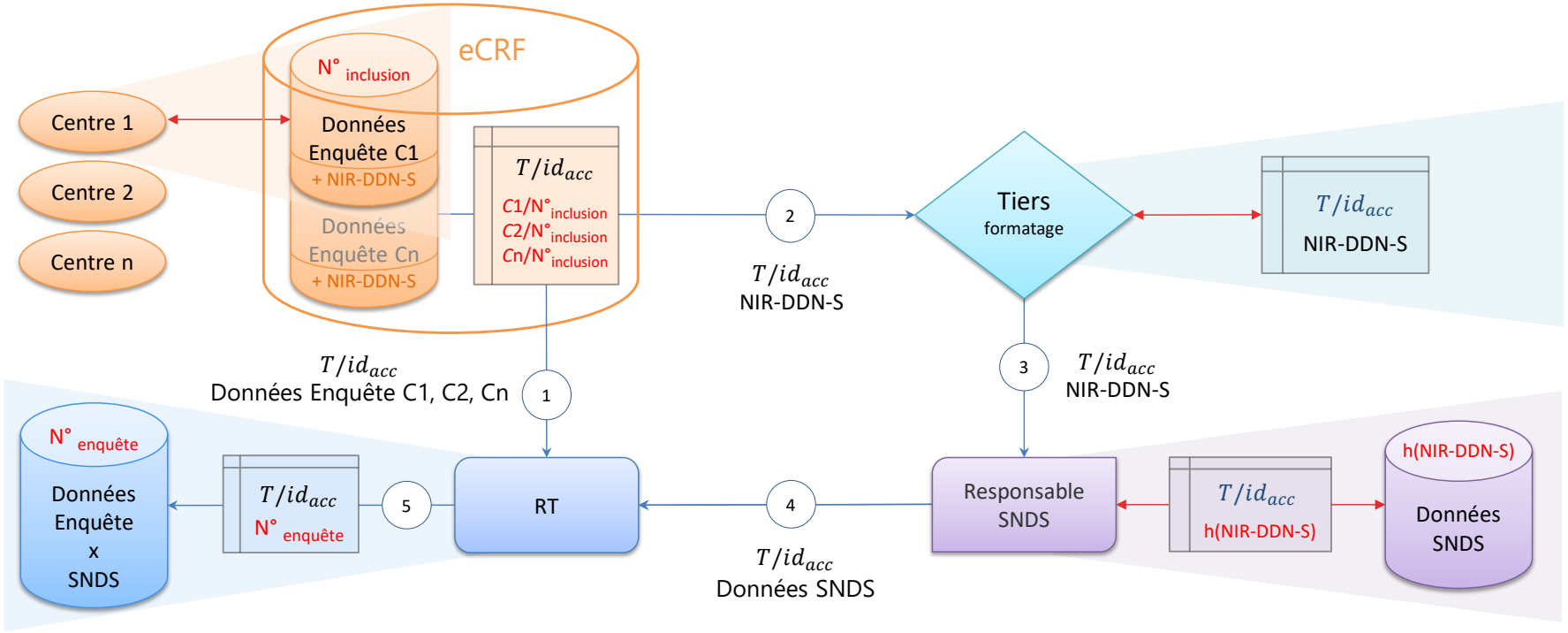
- Si le besoin est dûment justifié (par ex. transmission récurrente, audit des données, alerte des participants), la table de correspondance entre les identifiants d'accrochage et l'identifiant des données appariées peut être conservée par le responsable de traitement, de manière sécurisée.
- L'identifiant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



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

3. Synthèse de l'implémentation du circuit Multi-centres / eCRF avec NIR

Vue fonctionnelle complète



Vue technique complète