

# Guide pratique

MODALITES DE CIRCULATION DU NIR  
POUR LA RECHERCHE EN SANTE  
AUX FINS D'APPARIEMENT DE DONNEES  
AVEC LE SNDS

Décembre 2020

---

## 1. Problématiques identifiées

La CNIL encadre depuis de nombreuses années les appariements<sup>1</sup> de données de santé avec le SNIIRAM et, depuis sa création en 2016, avec le SNDS<sup>2</sup>. En raison des risques de réidentification, un principe général de cloisonnement des identifiants est appliqué et 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.

Dans ce cadre, différents types d'intermédiaires peuvent intervenir dans les circuits d'appariement et participer à leur sécurisation : tiers pour la centralisation des données, tiers pour la reconstruction du NIR à partir des traits d'identités et, parfois, un tiers pour la mise en forme des fichiers d'identités.

Plusieurs écueils fréquents ont été identifiés dans le cadre du traitement des demandes d'autorisation :

- le recours superflu à un tiers alors que le NIR est déjà connu du responsable de traitement ou du centre investigateur qui participe au projet de recherche ;
- une circulation inutile du NIR et/ou des données de santé ;
- un mauvais usage des identifiants techniques temporaires (appelés « numéros d'accrochage ») ;
- une conservation excessive de ces identifiants techniques, qui devraient être détruits rapidement ;
- l'implication d'une entité faisant en réalité partie du même organisme que le responsable de traitement et ne pouvant donc pas être considérée comme un « tiers ».

## 2. Recommandations générales concernant les tiers

Concernant le recours, lorsqu'il est nécessaire, à un tiers indépendant afin d'éviter que le responsable de traitement ne dispose des données identifiantes utilisées pour l'appariement (NIR ou traits d'identité des personnes), il convient de s'assurer que ce tiers est réellement indépendant du responsable de traitement et notamment :

- indépendance juridique : il s'agit d'une personne morale juridique distincte du responsable de traitement, il n'appartient pas au même organisme de recherche, il n'appartient pas au même centre de recherche, etc. ;
- indépendance économique : il n'est pas lié capitalistiquement (ni maison-mère, ni société-sœur, etc.), il n'est pas en situation de dépendance économique à l'égard du responsable de traitement (principal ou unique client), etc. ;
- il ne se trouve pas en situation de conflit d'intérêt vis-à-vis du responsable de traitement.

En outre, le tiers devra *a minima* respecter les mesures suivantes :

- o ne fournir ni le NIR ni les traits d'identité des personnes au responsable de traitement ;
- o n'associer le NIR avec aucune autre donnée que les données d'identification des personnes concernées (identifiants d'accrochage ou traits d'identité) ;
- o garantir la confidentialité de l'identifiant utilisé.

## 3. Schémas-types de circulation du NIR dans la recherche en santé

Concernant la circulation des données aux fins d'appariement, la CNIL a identifié des schémas-types représentant les principaux circuits conformes aux obligations légales. Les circuits présentés ici ont déjà été validés par la CNIL, soit dans le cadre d'avis rendus sur des projets de décrets en Conseil d'Etat, soit d'autorisations.

Les mesures techniques et organisationnelles nécessaires afin de sécuriser les échanges et stockage ne sont en revanche pas précisées ici, mais devront bien évidemment être mises en place afin de garantir la sécurité des données collectées<sup>3</sup>.

<sup>1</sup> Rapprochement d'ensembles de données distincts, à l'aide d'informations communes (par exemple, regroupement des données patient issues de différentes sources) ; [https://documentation-snds.health-data-hub.fr/fiches/appariement\\_donnees\\_externes.html#definition](https://documentation-snds.health-data-hub.fr/fiches/appariement_donnees_externes.html#definition)

<sup>2</sup> Dans le cadre d'avis sur des projets de décret puis d'autorisations, depuis l'entrée en vigueur de la LMSS de 2016, lorsque l'appariement nécessite le traitement du NIR.

<sup>3</sup> Voir <https://www.cnil.fr/fr/principes-cles/guide-de-la-securite-des-donnees-personnelles>

On notera que pour des raisons de sécurité, les bases de données du SNDS sont pseudonymisées et ne sont pas référencées directement avec le NIR mais avec un « NIR haché », c'est-à-dire ayant fait l'objet d'un traitement cryptographique irréversible et secret<sup>4</sup>. Il est noté « hNIR » dans les schémas ci-dessous. Pour permettre leur appariement avec le SNDS, les NIR des personnes d'une base externe au SNDS seront donc hachés de la même manière.

Comme ce hachage repose sur des secrets connus seulement des responsables des bases SNDS, il est nécessaire de leur transmettre, de manière très sécurisée, la liste des NIR des personnes sélectionnées. Dans le cas de la CNAM, qui opère la base centrale du SNDS et gère la fonction de hachage appelée « FOIN<sup>5</sup> », cette transmission sécurisée se fait sur un serveur dédié à travers une connexion chiffrée (« procédure SAFE »)<sup>6</sup>.

On notera également l'utilisation d'identifiants techniques temporaires (« identifiants d'accrochage ») permettant de dissocier le NIR et les données de santé lors des transferts entre acteurs. De préférence, ces numéros seront non significatifs et différents du numéro d'inclusion de la personne dans l'étude.

Ces transmissions devront être réalisées par des moyens de communication sécurisés, notamment à l'aide de mesures d'authentification et de chiffrement à l'état de l'art<sup>7</sup>.

Enfin, la manipulation des données appariées devra s'effectuer dans un environnement homologué conformément au référentiel de sécurité du SNDS<sup>8</sup>.

### 3.1 NIR détenu par le responsable de traitement ou dans chaque centre investigateur

#### Circuit « responsable de traitement » (RT)

Lorsque le responsable de traitement souhaite appairer ses données avec les données du SNDS et a déjà connaissance du NIR, le circuit est le plus simple :

1. le responsable de traitement envoie au responsable de la base SNDS un identifiant d'accrochage<sup>9</sup> ( $id_{acc}$ ) associé au NIR de chaque patient ;
2. le responsable de la base SNDS identifie en conséquence les données pertinentes au sein du SNDS et les transmet au responsable de traitement avec l'identifiant d'accrochage. Le NIR n'est jamais envoyé en retour pour des questions de sécurité (en particulier, le NIR ne doit jamais être rattaché aux données du SNDS). Le responsable de traitement dispose alors des données du SNDS qu'il pourra appairer grâce à l'identifiant d'accrochage et les conserver dans un environnement homologué, distinct du lieu de stockage du NIR.

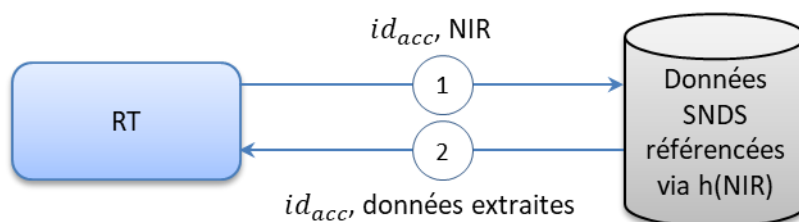


Figure 1 : NIR collecté par le responsable de traitement

<sup>4</sup> Voir <https://www.cnil.fr/fr/comprendre-les-grands-principes-de-la-cryptologie-et-du-chiffrement>

<sup>5</sup> FOIN : « Fonction d'occultation des identifiants nominatifs »

<sup>6</sup> Voir [https://documentation-snds.health-data-hub.fr/ressources/documents%20Cnam/guides%20p%C3%A9dagogiques%20SNDS/guide\\_pedagogique\\_acces\\_projets.html#quelles-sont-les-etapes-dans-l-extraction-des-donnees-du-snds](https://documentation-snds.health-data-hub.fr/ressources/documents%20Cnam/guides%20p%C3%A9dagogiques%20SNDS/guide_pedagogique_acces_projets.html#quelles-sont-les-etapes-dans-l-extraction-des-donnees-du-snds)

<sup>7</sup> Dans ce cadre, les outils cryptographiques utilisés devraient être conformes à l'annexe B1 du RGS.

<sup>8</sup> Voir <https://www.snds.gouv.fr/SNDS/Protection-de-la-donnee>

<sup>9</sup> L'identifiant d'accrochage est un identifiant individuel non significatif, de préférence différent du numéro d'inclusion de la personne dans l'étude

### **Circuit « centre investigateur unique » (CU)**

Lorsque l'étude implique que des données d'enquête et le NIR sont recueillis par un centre investigateur et que le responsable de traitement n'a pas l'autorisation de traiter le NIR, le circuit doit évoluer :

1. d'une part, le centre envoie les données d'enquête au responsable de traitement avec un identifiant d'accrochage ( $id_{acc}$ );
2. d'autre part, le centre transmet au responsable de la base SNDS ce même identifiant d'accrochage avec le NIR ;
3. le responsable de la base SNDS identifie en conséquence les données pertinentes au sein du SNDS et les transmet au responsable de traitement avec l'identifiant d'accrochage. Le responsable de traitement peut alors appairer ensemble les données d'enquête et les données issues du SNDS grâce à l'identifiant d'accrochage.

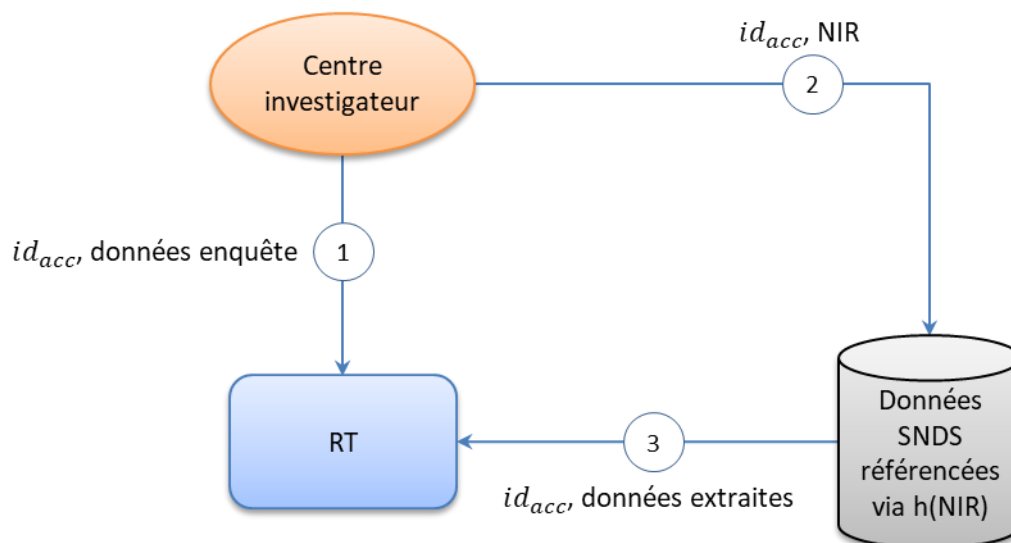


Figure 2 : NIR collecté par le centre investigateur

### **Circuit « multi-centres » (MC)**

Lorsque la recherche est menée dans plusieurs centres investigateurs, chaque centre collecte les données d'enquête ainsi que le NIR pour ses participants ; le recours à un tiers est alors indispensable afin de centraliser l'ensemble des données nécessaires pour l'appariement :

1. chaque centre envoie au responsable de traitement les données d'enquête associées à des identifiants d'accrochage générés au sein de chaque centre ( $C1/id_{acc}$  pour le premier centre, etc.) ;
2. d'autre part, chacun des centres transmet ce même identifiant d'accrochage avec le NIR au tiers de centralisation ;
3. le tiers est alors capable d'envoyer au responsable de la base SNDS, dans le format adapté et en une unique fois, les identifiants d'accrochage associés aux NIR des participants ;
4. le responsable de la base SNDS transmet les données issues du SNDS au responsable de traitement avec l'identifiant d'accrochage.

NB : Si certaines personnes sont rattachées à deux centres différents, elles auront deux identifiants différents pour l'étude (par ex.  $C1/id_{acc}$  et  $C3/id_{acc}$ ). Dans ce cas, le tiers de centralisation va dédoubler / fusionner les données et générer un nouvel identifiant d'accrochage ( $T/id_{acc}$ ). Il doit alors transmettre au responsable de traitement la table de correspondance entre les identifiants des centres et ce nouvel identifiant, afin de permettre à ce dernier d'appairer les données qu'il a reçues des centres avec celles du SNDS.

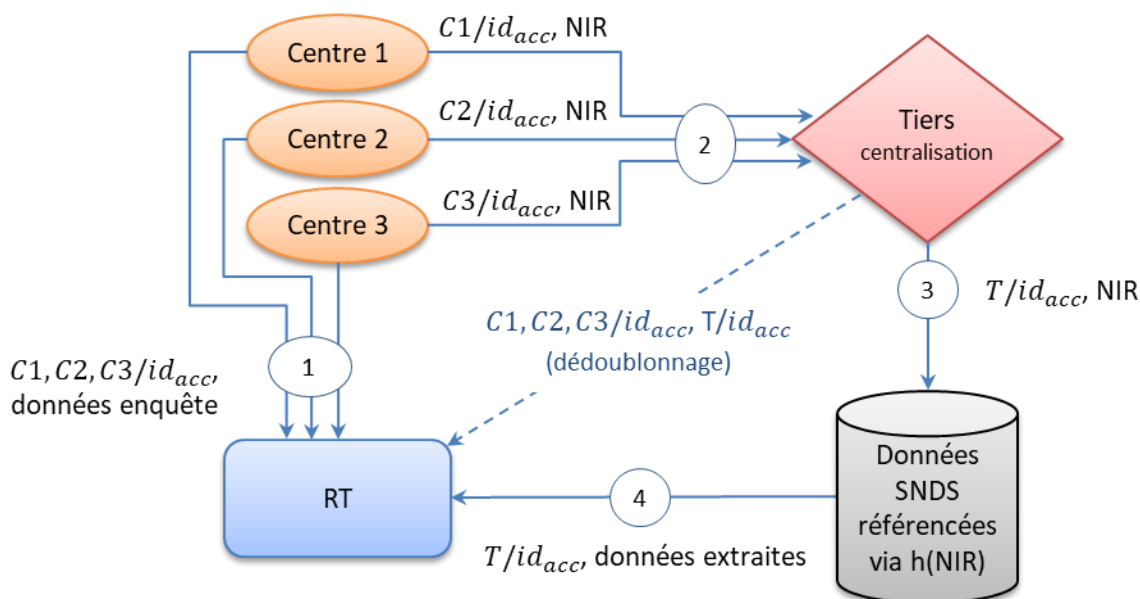


Figure 3: Recherche multicentrique - NIR centralisé par un tiers

## 3.2 NIR reconstitué par un tiers

### ***Circuit « reconstruction du NIR » (RN)***

Dans certaines études, recherches ou évaluations, le NIR n'est pas directement collecté mais est néanmoins nécessaire pour permettre l'appariement. Dans ce cas, il doit être reconstitué à partir des traits d'identités (nom, prénoms, sexe, date et lieu de naissance) des personnes concernées par l'intermédiaire du RNIPP<sup>10</sup> ou du RNIAM<sup>11</sup>. Le rôle de tiers de reconstruction est alors attribué à la CNAV ou à l'INSEE. Les différents schémas précédents doivent alors être mis à jour avec un nouveau tiers dit de « reconstruction ».

Lorsque des données d'enquête sans le NIR sont recueillies par un centre investigateur, le circuit devient alors le suivant :

1. d'une part, le centre envoie les données d'enquête au responsable de traitement avec un identifiant d'accrochage ( $id_{acc}$ ) ;
2. d'autre part, le centre transmet au tiers de reconstruction les traits d'identités<sup>12</sup> (traits) avec ce même identifiant d'accrochage ;
3. le tiers reconstruit le NIR à partir des traits d'identité et le transmet alors au responsable de la base SNDS avec l'identifiant d'accrochage ;
4. comme dans les cas précédents, le responsable de la base SNDS identifie en conséquence les données pertinentes au sein du SNDS et les transmet au responsable de traitement avec l'identifiant d'accrochage. Le responsable de traitement peut alors appairer les données d'enquête et les données issues du SNDS grâce à l'identifiant d'accrochage.

<sup>10</sup> Répertoire national d'identification des personnes physiques

<sup>11</sup> Répertoire national interrégimes des bénéficiaires de l'assurance maladie

<sup>12</sup> Nom, prénoms, sexe, date et lieu de naissance

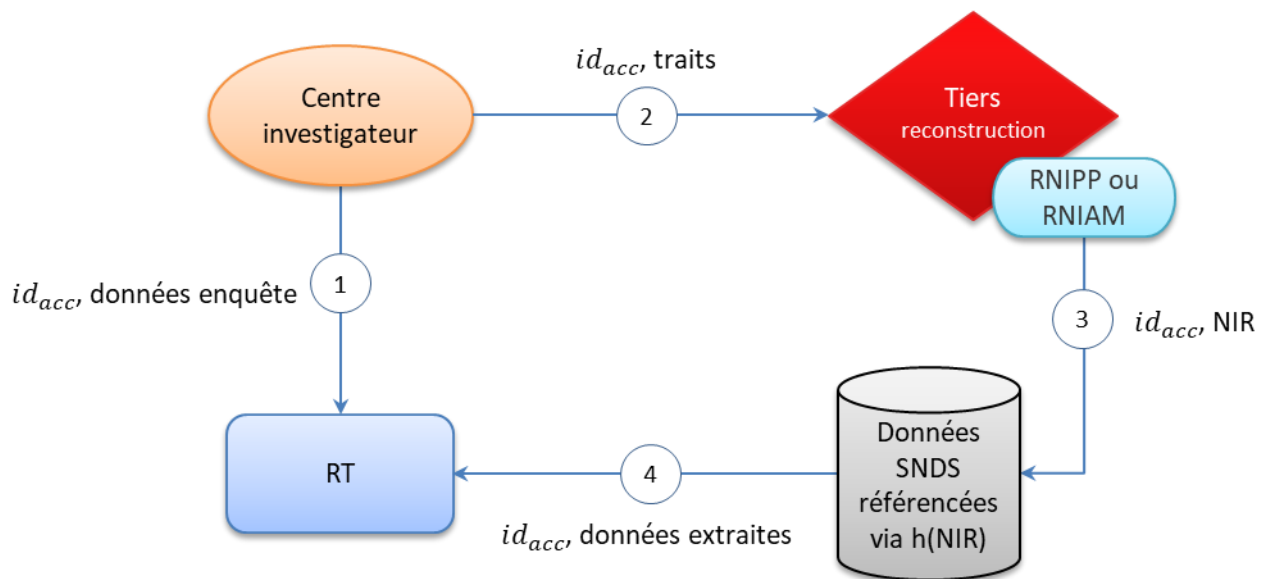


Figure 4 : NIR reconstitué par un tiers

### Circuit « multi-centres et reconstruction du NIR » (MC+RN)

Dans le cas d'une recherche dans plusieurs centres investigateurs, le tiers de reconstruction du NIR intervient après le tiers de centralisation qui est chargé de réunir les traits d'identités (traits) et identifiants d'accrochage ( $C1/id_{acc}$  pour le premier centre, etc.) issus de l'ensemble des centres.

NB : Comme pour le circuit R3, le tiers de centralisation peut être amené à générer un nouvel identifiant d'accrochage ( $T/id_{acc}$ ) après avoir dédoublonné les personnes présentes dans plusieurs centres.

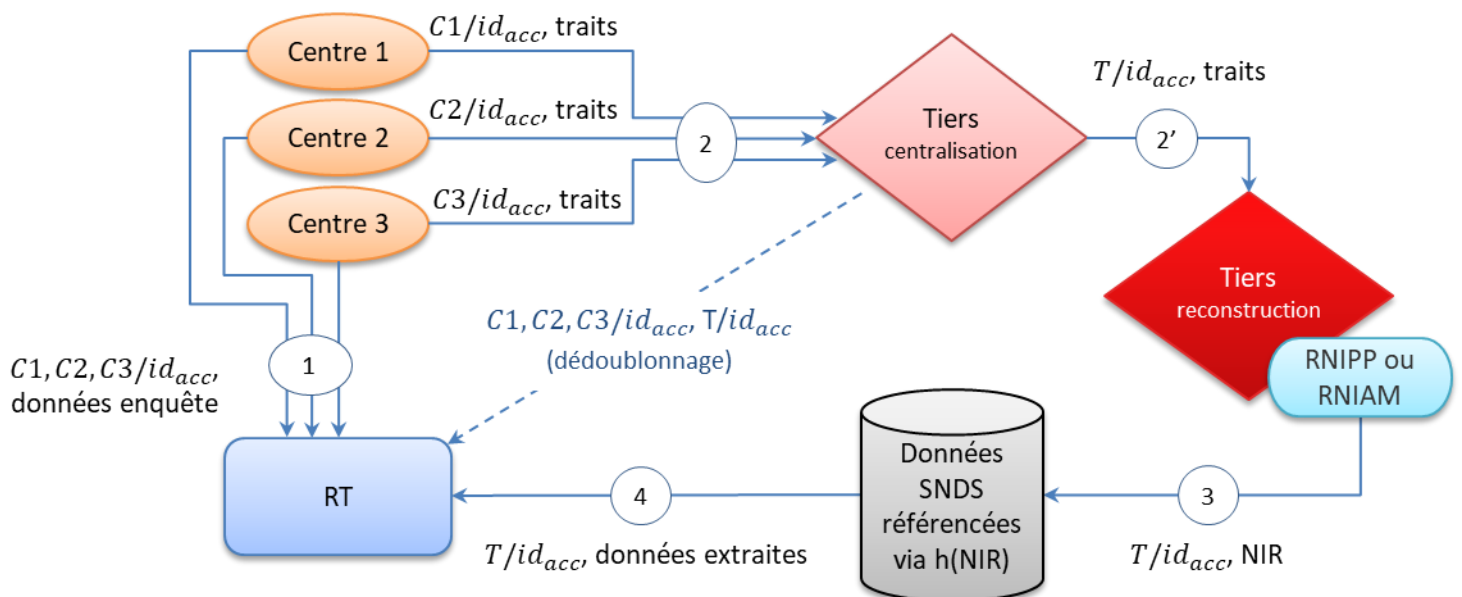


Figure 5 : Recherche multicentrique - NIR centralisé par un tiers, puis reconstitué par un autre tiers

## 4. Choix du recours à un tiers

Le recours à un tiers n'a pas à être systématique. Dans certains cas, le NIR peut être transmis directement au responsable de la base SNDS par les centres investigateurs après un recueil direct auprès des personnes concernées.

Mais dans d'autres cas, par exemple en raison de contraintes techniques liées au format des données fixé par la CNAM pour la procédure « SAFE », il peut être nécessaire qu'un tiers mette préalablement les informations dans le format adéquat. Cela ne concerne alors que le NIR et l'identifiant d'accrochage. Le tiers n'a aucune connaissance des données associées (celles de l'enquête comme celles du SNDS).

On ajoute alors une étape supplémentaire de formatage, illustrée ci-dessous sur le circuit R1 :

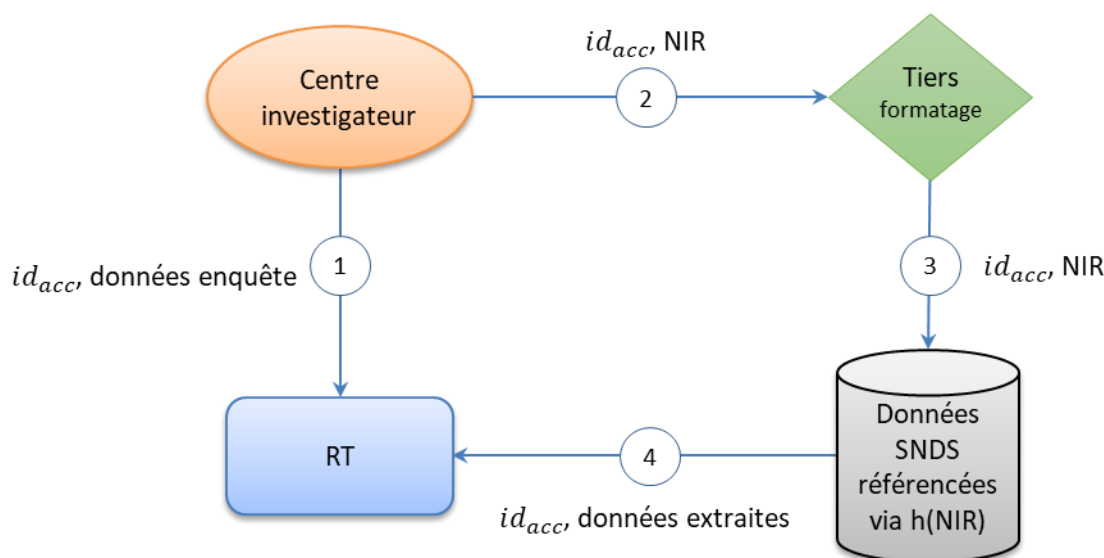


Figure 6 - Cas d'une mise en forme préalable des données réalisée par un tiers

- On notera que quel que soit le type de prestation réalisée par le tiers (formatage, centralisation, reconstruction), il ne réalise jamais d'appariement entre des données d'identification et des données relatives à la santé. Il ne réalise que des appariements d'identité pour le compte du responsable de traitement, entre plusieurs identifiants des personnes concernées (numéro d'accrochage, traits d'identité, NIR, etc.).
- On notera également que, afin de respecter le principe général de cloisonnement des identifiants visant à limiter les risques de réidentification des données, chaque acteur des différents circuits présentés dans ce guide ne peut recevoir que les données strictement nécessaires à son action, telles qu'indiquées dans les schémas correspondants.

Pour ces différentes raisons, il sera nécessaire de déterminer au préalable la pertinence du recours à un tiers pour un appariement de données mobilisant le NIR. Les critères suivants peuvent alors être pris en compte :

**1. S'agit-il d'une recherche multicentrique ?**

Si oui, un tiers de centralisation est alors requis.

**2. Le NIR est-il collecté à la source ?**

Si oui, l'envoi du NIR et du numéro d'accrochage à la CNAM peut être réalisé directement par le responsable investigateur ou le centre : un tiers est inutile.

Si non, un tiers de reconstruction est nécessaire.

Le tableau ainsi que l'arbre de décision ci-après permettent de déterminer facilement la pertinence du recours aux tiers de centralisation et de reconstruction. La nécessité du recours au tiers de formatage dépend quant à lui du contexte et des moyens techniques utilisés ; il n'est donc pas mentionné explicitement sur tous les schémas mais peut être sollicité dans chaque cas si aucun autre moyen n'a pu être trouvé.

	<b>NIR collecté</b>	<b>NIR non collecté</b>
Recherche dans un unique centre investigateur	aucun tiers nécessaire (schéma RT ou CU)	tiers de reconstruction (schéma RN)
Recherche dans plusieurs centres investigateurs	tiers de centralisation (schéma MC)	tiers de centralisation + tiers de reconstruction (schéma MC+RN)

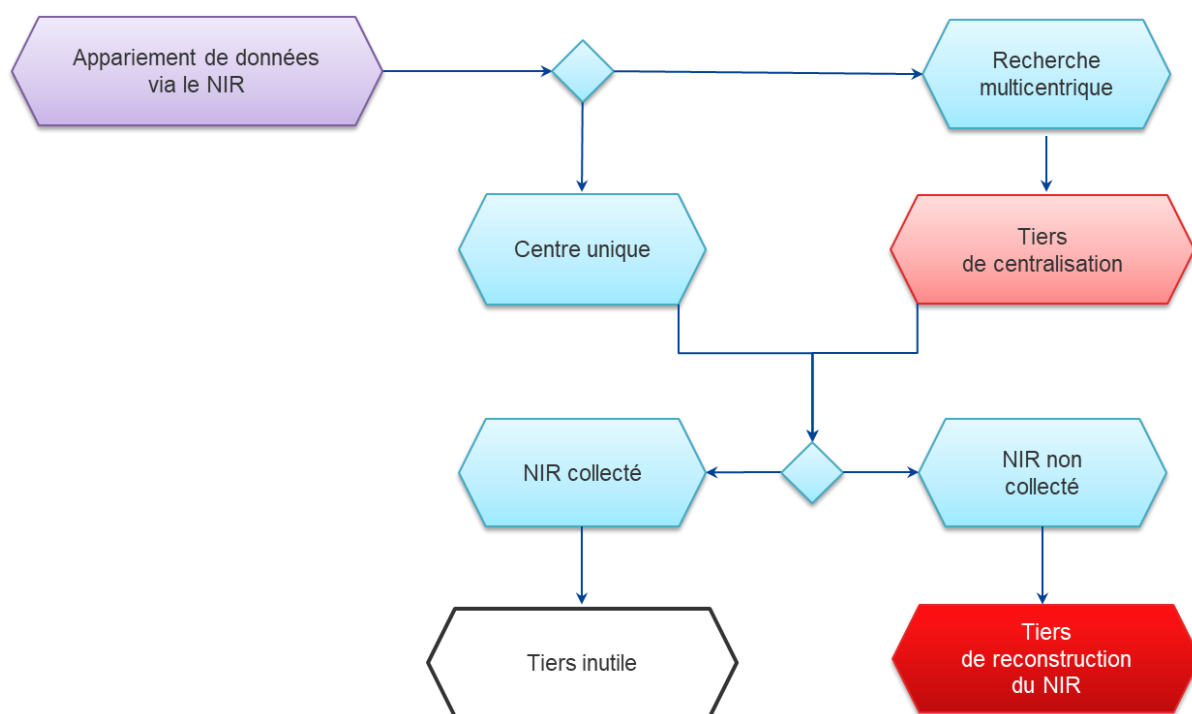


Figure 7 : Arbre de décision pour le recours à un tiers