



Qualifier une donnée à caractère personnel en santé

Sorbonne Université, Faculté de Santé

1.2 Caractériser et traiter la donnée à caractère personnel de santé en appliquant la réglementation

CC-BY-NC 4.0

02/04/2026

données caractère personnel, données sensibles, données concernant la santé
données directement identifiantes, données non directement identifiantes,
pseudonymisation
données anonymisées
données agrégées

Qualifier une donnée à caractère personnel en santé

Pr Ferdinand DHOMBRES

Sorbonne Université, Faculté de Santé



OBJECTIFS PÉDAGOGIQUES

Connaitre les **définitions** et le **vocabulaire** liés aux données personnelles

Savoir **qualifier** une donnée de santé

Connaitre les différents **niveaux de données identifiantes**



Quelle est la valeur des données personnelles ?

Un exemple (parmi d'autres) :

- **Affaire Cambridge Analytica**
- Lanceur d'alerte Christopher Wylie
- Société spécialisée de développement d'outils d'influence des électeurs
- Données de 30 à 50M d'utilisateurs du réseau social FB traitées sans consentement
- La campagne de Donald Trump (élection présidentielle de 2016) a versé 6M de dollars à cette entreprise...

Le Monde

The New York Times

2 Days, 10 Hours, 600 Questions: What Happened When Mark Zuckerberg Went to Washington

Give this article



Senator John Kennedy told Mark Zuckerberg, the chief executive of Facebook, that his company's user agreement "sucks." Our reporter Sheera Frenkel explains the senator's questions, Mr. Zuckerberg's answers and what they really mean.
Tom Brenner/The New York Times

10 avril 2018

Protéger ses données : réflexions autour de l'argument "je n'ai rien à cacher"

- Vos habitudes alimentaires

Quel(s) mésusage(s)
possible(s) ? Préjudice ?

Protéger ses données : réflexions autour de l'argument "je n'ai rien à cacher"

- Vos habitudes alimentaires
- Vos données de santé

Quel(s) mésusage(s)
possible(s) ? Préjudice ?

Protéger ses données : réflexions autour de l'argument "je n'ai rien à cacher"

- Vos habitudes alimentaires
- Vos données de santé



Coût d'une assurance / complémentaire santé ?



Obtention d'un prêt sur 20 ans ?

Protéger ses données : réflexions autour de l'argument "je n'ai rien à cacher"

- Vos données de santé
- Vos habitudes alimentaires
- Vos données de navigation, d'achat en ligne
- Vos relations sur les réseaux sociaux
- Vos déplacements
- ... ,

et bien entendu vos identifiants / mots de passe pour y accéder

Toutes ces données sensibles ont de la valeur et certaines se "monétisent" (plutôt bien) sur le dark web !

Quelle est la valeur des données personnelles sur le dark web ?

- Selon le *Dark Web Price Index 2023* (enquête sur les activités récentes de cybercriminalité de 2022 à Q1 2023) :

	Avg. dark web Price
Credit card details, account balance up to 5,000	\$110
Credit card details, account balance up to 1,000	\$70
Revolut verified account (UK, USA)	\$1,600
Hacked Gmail account	\$60
Hacked Facebook account	\$25
Hacked Instagram account	\$25
Hacked Twitter account	\$20
(...)	

PLAN



DONNÉES À CARACTÈRE PERSONNEL : QUELS ENJEUX ?

Cas de Cambridge Analytica



DÉFINITIONS

Données à caractère personnel, données sensibles, données de santé



CARACTERE IDENTIFIANT D'UNE DONNEE

données directement identifiantes, données non directement identifiantes, données anonymisées, données agrégées

LES DONNÉES À CARACTÈRE PERSONNEL

Exemples et définition



Quelles sont les données manipulées par les professionnels de santé ?

- **informations pour gérer un cabinet/structure**
 - (ex : gestion des fournisseurs, des personnels, etc.)
- **informations sur les patients**
 - données d'identification
 - (ex : nom, prénom, adresse, numéro de téléphone...)
 - informations sur la vie personnelle du patient
 - (ex : nombre d'enfants, profession...)
 - informations sur la couverture sociale
 - (ex : assurance maladie obligatoire, complémentaire...)
 - informations relatives à sa santé
 - (ex : pathologie, diagnostic, prescriptions, soins, etc.)
 - éventuels professionnels/correspondants
 - (ex : les intervenants dans la prise en charge du patient)
 - numéro de sécurité sociale des patients
 - (Numéro d'Inscription au Répertoire des Personnes Physiques - NIR) pour la facturation



Image by Freepik

Quelles sont les données manipulées par les professionnels de santé ?

- informations pour gérer un cabinet/structure
 - (ex : gestion des fournisseurs, des personnels, etc.)
- informations sur les patients
 - données d'identification
 - (ex : nom, prénom, adresse, numéro de téléphone, etc.)
 - informations sur la vie personnelle du patient
 - (ex : nombre d'enfants, profession, etc.)
 - informations sur la couverture sociale
 - (ex : assurance maladie obligatoire, complémentaire, etc.)



Image by Freepik

Ces informations que vous recevez et/ou émettez, à l'occasion de votre activité professionnelle, sont considérées comme des **données à caractère personnel**.

«données à caractère personnel»

- toute information se rapportant à une personne physique identifiée ou identifiable
- personne physique identifiable = une personne physique qui peut être identifiée, **directement ou indirectement**, notamment par référence à
 - un identifiant, tel qu'un nom, un numéro d'identification, des données de localisation, un identifiant en ligne
 - un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, génétique, psychique, économique, culturelle ou sociale

«données à caractère personnel»

- toute information se rapportant à une personne physique identifiée ou identifiable

directement ou indirectement

Numéro de téléphone
Adresse mail
Numéro d'abonnement
...



«données à caractère personnel»

- **toute information se rapportant à une personne physique identifiée ou identifiable**

Qu'elle soit **confidentielle ou publique**, de nature **privée ou professionnelle**, toute information qui correspond à cette définition est considérée comme une donnée à caractère personnel.



LES DONNÉES SENSIBLES, LES DONNÉES CONCERNANT LA SANTÉ



Notion de donnée “sensible”

<https://www.cnil.fr/fr/definition/donnee-sensible>

- **catégorie particulière des données personnelles**
- correspondant aux informations qui révèlent :
 - la prétendue origine raciale ou ethnique,
 - les opinions politiques,
 - les convictions religieuses ou philosophiques ou l'appartenance syndicale,
- ainsi que le traitement
 - **des données concernant la santé,**
 - des données génétiques,
 - des données biométriques aux fins d'identifier une personne physique de manière unique,
 - des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique.

Notion de donnée “sensible”

<https://www.cnil.fr/fr/definition/donnee-sensible>

- catégorie particulière des données personnelles
- **Le RGPD interdit le recueil ou l'utilisation de ces données**, sauf :
 - si la personne concernée a donné son consentement exprès (démarche active, explicite et de préférence écrite, qui doit être libre, spécifique, et informée) ;
 - si les informations sont manifestement rendues publiques par la personne concernée ;
 - si elles sont nécessaires à la sauvegarde de la vie humaine ;
 - si leur utilisation est justifiée par l'intérêt public et autorisée par la CNIL ;
 - si elles concernent les membres ou adhérents d'une association ou d'une organisation politique, religieuse, philosophique, politique ou syndicale.

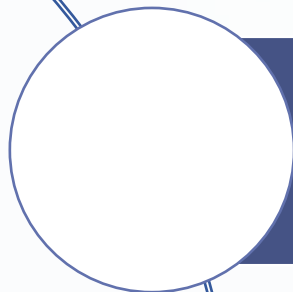
Les "données concernant la santé"

<https://www.cnil.fr/fr/definition/donnee-sensible>

- catégorie particulière des données personnelles
- relatives à la santé physique ou mentale d'une personne physique, y compris la prestation de services de soins de santé, qui révèlent des informations sur l'état de santé de cette personne

Données particulières car considérées comme sensibles.
→ protection particulière par les textes réglementaires
(RGPD, CNIL, etc.)

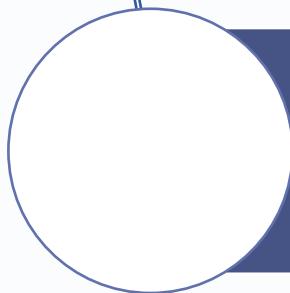
Catégories de "données concernant la santé"



Données de santé

PAR NATURE

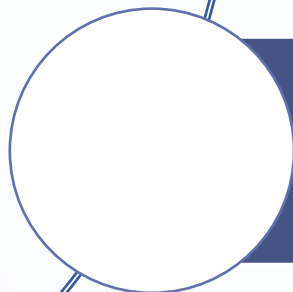
ATCD médicaux, diagnostics, prestations de soins réalisés, résultats d'examens, traitements, handicap...



Données de santé

PAR COMBINAISON

Du fait de leur croisement avec d'autres données, elles deviennent des données de santé (ex : poids + nombre de pas)



Données de santé

PAR DESTINATION

De part l'utilisation qui en est faite au plan médical, elles deviennent des données de santé (ex: une photo)

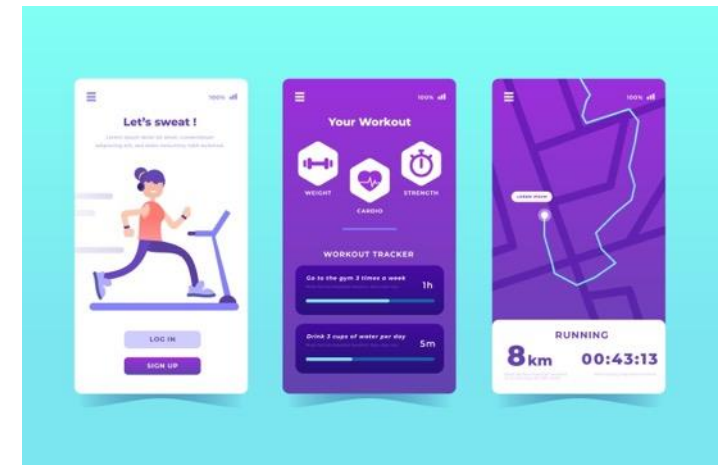
Précisions sur les «données concernant la santé»

- **N'entrent pas dans la notion de données de santé celles à partir desquelles aucune conséquence ne peut être tirée au regard de l'état de santé de la personne concernée.**

ex : une application collectant un nombre de pas au cours d'une promenade sans croisement de ces données avec d'autres.

- **La loi ne s'applique pas aux traitements qui comporteraient des données de santé à l'usage exclusif de la personne.**

ex : les applications mobiles en santé avec collecte, enregistrement ou conservation de données, à condition que ces opérations s'effectuent localement, sans connexion extérieure et à des fins exclusivement personnelles.



«données génétiques»



ARTICLE 4 - DÉFINITIONS

- **données à caractère personnel**
- « relatives aux caractéristiques génétiques héréditaires ou acquises d'une personne physique qui donnent des informations uniques sur la physiologie ou l'état de santé de cette personne physique et qui résultent, notamment, d'une analyse d'un échantillon biologique de la personne physique en question »



«données biométriques»



ARTICLE 4 - DÉFINITIONS

- **données à caractère personnel**
- « résultant d'un traitement technique spécifique, relatives aux caractéristiques physiques, physiologiques ou comportementales d'une personne physique, qui permettent ou confirment son identification unique, telles que des images faciales ou des données dactyloscopiques »



Parmi ces données, laquelle ou lesquelles ne sont pas des données personnelles ?

Un numéro de plaque d'immatriculation

Des coordonnées d'entreprise

Un numéro de carte de paiement

Des images de vidéosurveillance

Parmi ces données, laquelle ou lesquelles ne sont pas des données personnelles ?

Un numéro de plaque
d'immatriculation

Des coordonnées
d'entreprise

Un numéro
de carte de paiement

Des images
de vidéosurveillance

PLAN



DONNÉES À CARACTÈRE PERSONNEL : QUELS ENJEUX ?

Cas de Cambridge Analytica



DÉFINITIONS

Données à caractère personnel, données sensibles, données de santé



CARACTERE IDENTIFIANT D'UNE DONNEE

données directement identifiantes, données non directement identifiantes, données anonymisées, données agrégées

LE CARACTÈRE IDENTIFIANT DES DIFFÉRENTS TYPES DE DONNÉES



Les “niveaux” d’identification des données

Niveau fort



- **Données identifiantes**

- Directement identifiantes :
 - données nominatives (nom, prénom), ...
- Non directement identifiantes :
 - DDN, dates de RDV, ...
 - données codées / pseudonymisées

- **Données anonymisées**

- **Données agrégées**

Niveau faible

Notion de « pseudonymisation »

« traitement de données à caractère personnel de telle façon que celles-ci **ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires**, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable »

- La pseudonymisation constitue une des mesures recommandées par le RGPD pour limiter les risques liés au traitement de données personnelles.
- Elle est souvent utilisée dans le domaine de la recherche en santé.

Notion de « pseudonymisation »

« traitement de données à caractère personnel de telle façon que celles-ci ne puissent plus être attribuées à une personne concernée précise sans avoir recours à des informations supplémentaires, pour autant que ces informations supplémentaires soient conservées séparément et soumises à des mesures techniques et organisationnelles afin de garantir que les données à caractère personnel ne sont pas attribuées à une personne physique identifiée ou identifiable »

- La pseudonymisation constitue une des **mesures recommandées par le RGPD pour limiter les risques liés au traitement de données personnelles**.
- Elle est souvent utilisée dans le domaine de la recherche en santé.

Notion de « pseudonymisation » : EXEMPLE

La pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.)

Nom	Prénom	DDN	Diagnostic
aaa	abab	1/2/77			COVID
bbb	bcbc	4/5/83			BPCO
...	...				
zzz	zaazaa	8/9/81			ACFA

Données nominatives

Id	ADN	Diagnostic
001	1977			COVID
002	1983			BPCO
...				
999	1981			ACFA

Données pseudonymisées

Notion de « pseudonymisation » : EXEMPLE

La pseudonymisation consiste à remplacer les données directement identifiantes (nom, prénom, etc.) d'un jeu de données par des données indirectement identifiantes (alias, numéro séquentiel, etc.)

Nom	Prénom	DDN	Diagnostic
aaa	abab	1/2/77			COVID
bbb	bcbc	4/5/83			
...	...				
zzz	zaazaa	8/9/81			

Données nominatives

Id	Nom	Prénom	DDN
001	aaa	abab	1/2/77
002	bbb	bcbc	4/5/83
...	
999	zzz	zaazaa	8/9/81

Table de correspondance

Id	ADN	Diagnostic
001	1977			COVID
	1983			BPCO
	1981			ACFA

Données pseudonymisées

Notion de « donnée anonyme »

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre **impossible**, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière **irréversible**.



Notion de « donnée anonyme »

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre **impossible**, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière **irréversible**.

CONSÉQUENCES D'UNE ANONYMISATION DE DONNÉES

- la législation relative à la protection des données ne s'y applique plus
- elle permet d'exploiter des données personnelles dans le respect des droits et libertés des personnes
- elle ouvre des potentiels de réutilisation des données initialement interdits
- elle permet de conserver des données sans limite de temps

Notion de « donnée anonyme »

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre **impossible**, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière **irréversible**.

La mise en œuvre d'un procédé d'anonymisation de données nécessite généralement :

- d'identifier les informations à conserver selon leur pertinence
- de supprimer les éléments d'identification directe ainsi que les valeurs rares qui pourraient permettre une ré-identification aisée des personnes (ex : âge / centenaires)
- de sélectionner les informations importantes (et supprimer les autres)
- de définir la finesse idéale et acceptable pour chaque information conservée

Notion de « donnée anonyme »

L'anonymisation est un traitement qui consiste à utiliser un ensemble de techniques de manière à rendre **impossible**, en pratique, toute identification de la personne par quelque moyen que ce soit et de manière **irréversible**.

Les deux principales familles de techniques d'anonymisation sont :

- **la randomisation** : modifier les attributs dans un jeu de données de telle sorte qu'elles soient moins précises, tout en conservant la répartition globale.
- **la généralisation** : modifier l'échelle des attributs des jeux de données, ou leur ordre de grandeur, afin de s'assurer qu'ils soient communs à un ensemble de personnes.

Notion de « donnée anonyme » : critères d'évaluation

Le « groupe de travail de l'article 29 » (qui regroupe les autorités de protection des données européennes) a publié un avis sur les principales techniques d'anonymisation et propose trois critères suivants pour évaluer une solution d'anonymisation :

1. **L'individualisation** : est-il toujours possible d'isoler un individu ?
2. **La corrélation** : est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
3. **L'inférence** : peut-on déduire de l'information sur un individu ?

GRUPE DE TRAVAIL «ARTICLE 29» SUR LA PROTECTION DES
DONNÉES



0829/14/FR
WP216



Notion de « donnée anonyme » : critères d'évaluation

Le « groupe de travail de l'article 29 » (qui regroupe les autorités de protection des données européennes) a publié un avis sur les principales techniques d'anonymisation et propose trois critères suivants pour évaluer une solution d'anonymisation :

1. **L'individualisation** : est-il toujours possible d'isoler un individu ?
2. **La corrélation** : est-il possible de relier entre eux des ensembles de données distincts concernant un même individu ?
3. **L'inférence** : peut-on déduire de l'information sur un individu ?

La **randomisation** permet de protéger le jeu de données du risque **d'inférence**.
La **généralisation** permet d'éviter **l'individualisation** d'un jeu de données et limite les possibles **corrélations** du jeu de données avec d'autres



Notion de « donnée anonyme » : critères d'évaluation

Ainsi il faut retenir que :

1. un ensemble de données pour lequel il n'est possible ni d'individualiser ni de corréliser ni d'inférer est a priori anonyme.
2. un ensemble de données pour lequel au moins un des trois critères n'est pas respecté ne pourra être considéré comme anonyme qu'à la suite d'une analyse détaillée des risques de ré-identification.

individualisation

corrélation

inférence



Notion de « donnée anonyme »

EXEMPLE

il ne doit pas être possible d'isoler un individu dans le jeu de données

- *Exemple : une base de données de CV où seuls les nom et prénoms d'une personne auront été remplacés par un numéro (qui ne correspond qu'à elle) permet d'individualiser cette personne.*

Dans ce cas, cette base de données est considérée comme pseudonymisée et non comme anonymisée.

Notion de « donnée anonyme »

EXEMPLE

il ne doit pas être possible de relier entre eux des ensembles de données distincts concernant un même individu

- **Exemple** : *une base de données cartographiques renseignant les adresses des domiciles de particuliers ne peut être considérée comme anonyme si d'autres bases de données, existantes par ailleurs, contiennent ces mêmes adresses avec d'autres données permettant d'identifier les individus.*

Quelle solution proposer ?

Notion de « donnée anonyme »

EXEMPLE

il ne doit pas être possible de déduire, de façon quasi certaine, de nouvelles informations sur un individu.

- **Exemple** : *si un jeu de données supposément anonyme contient des informations sur le montant des impôts de personnes ayant répondu à un questionnaire, que tous les hommes ayant entre 20 et 25 ans qui ont répondu sont non imposables, il sera possible de déduire, si on sait que M. X, homme âgé de 24 ans, a répondu au questionnaire, que ce dernier est non imposable.*

Notion de « données agrégées »

- Ils s'agit de nouvelles données calculées à partir du jeu de données source, généralement des données de nature statistiques :
 - Moyenne
 - Médiale
 - Quartiles
 - ...

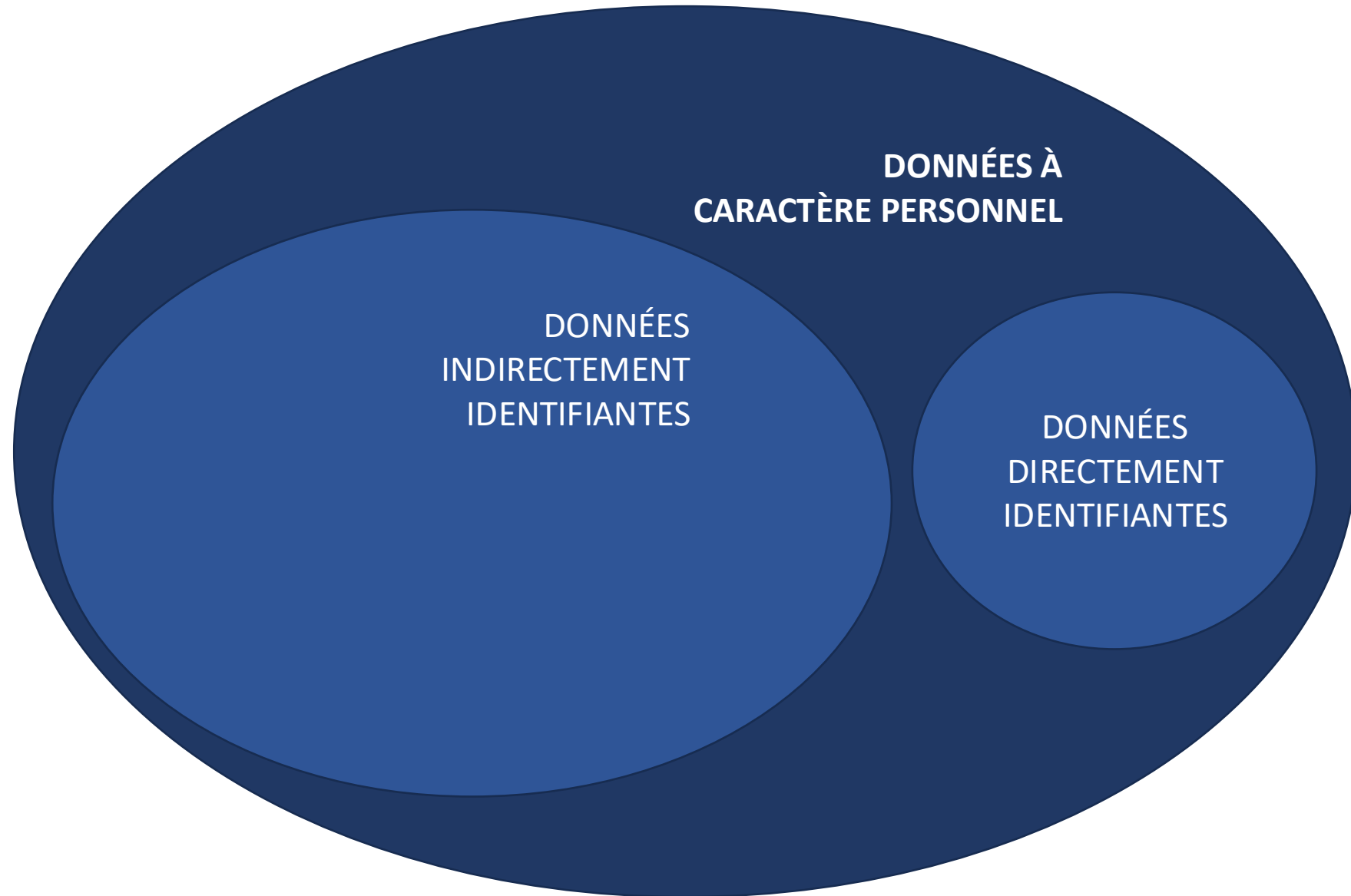
Ces données agrégées ne comportent généralement pas de données à caractère personnel (mais il peut y avoir des exceptions)

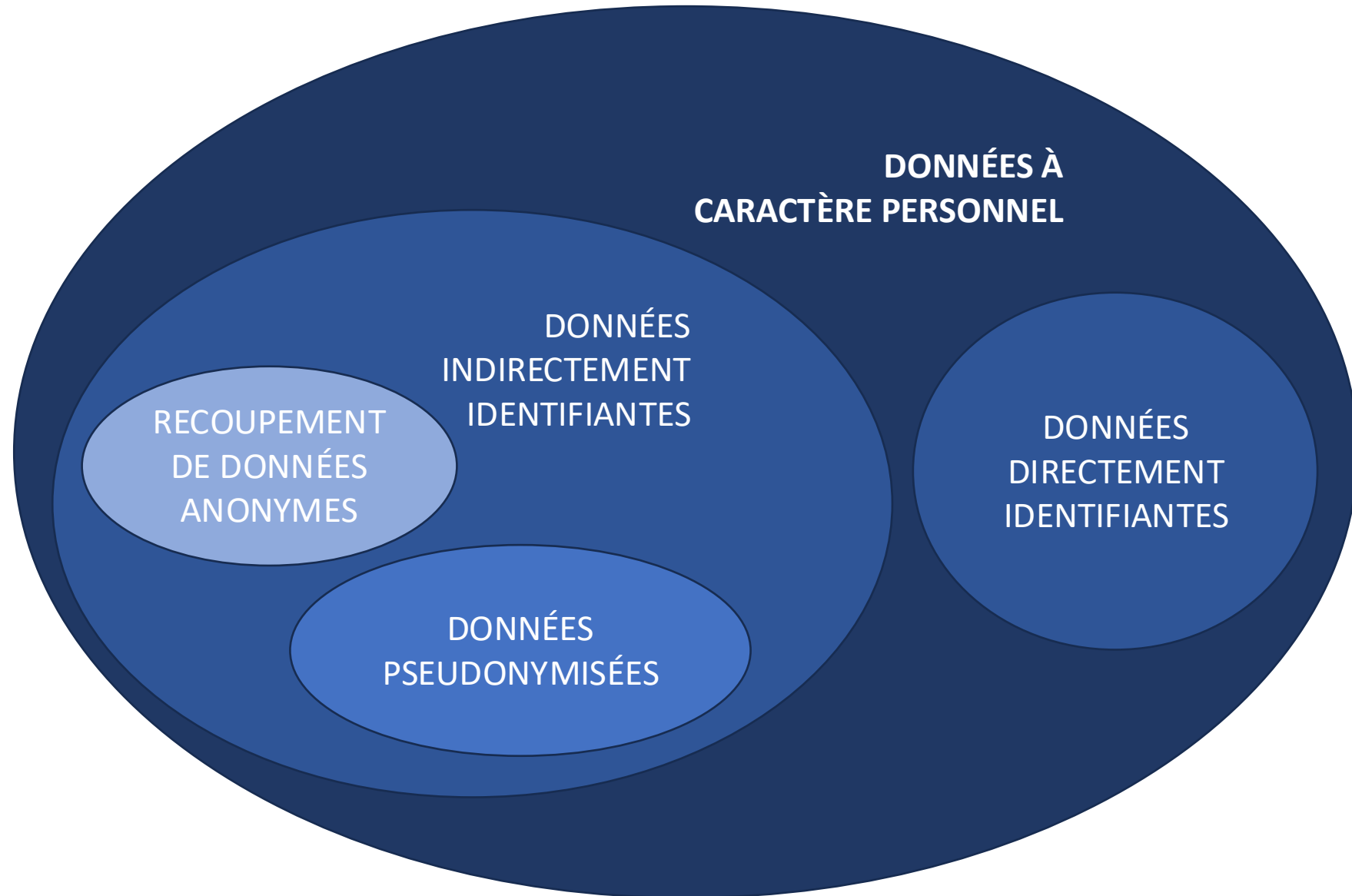
Pour résumer



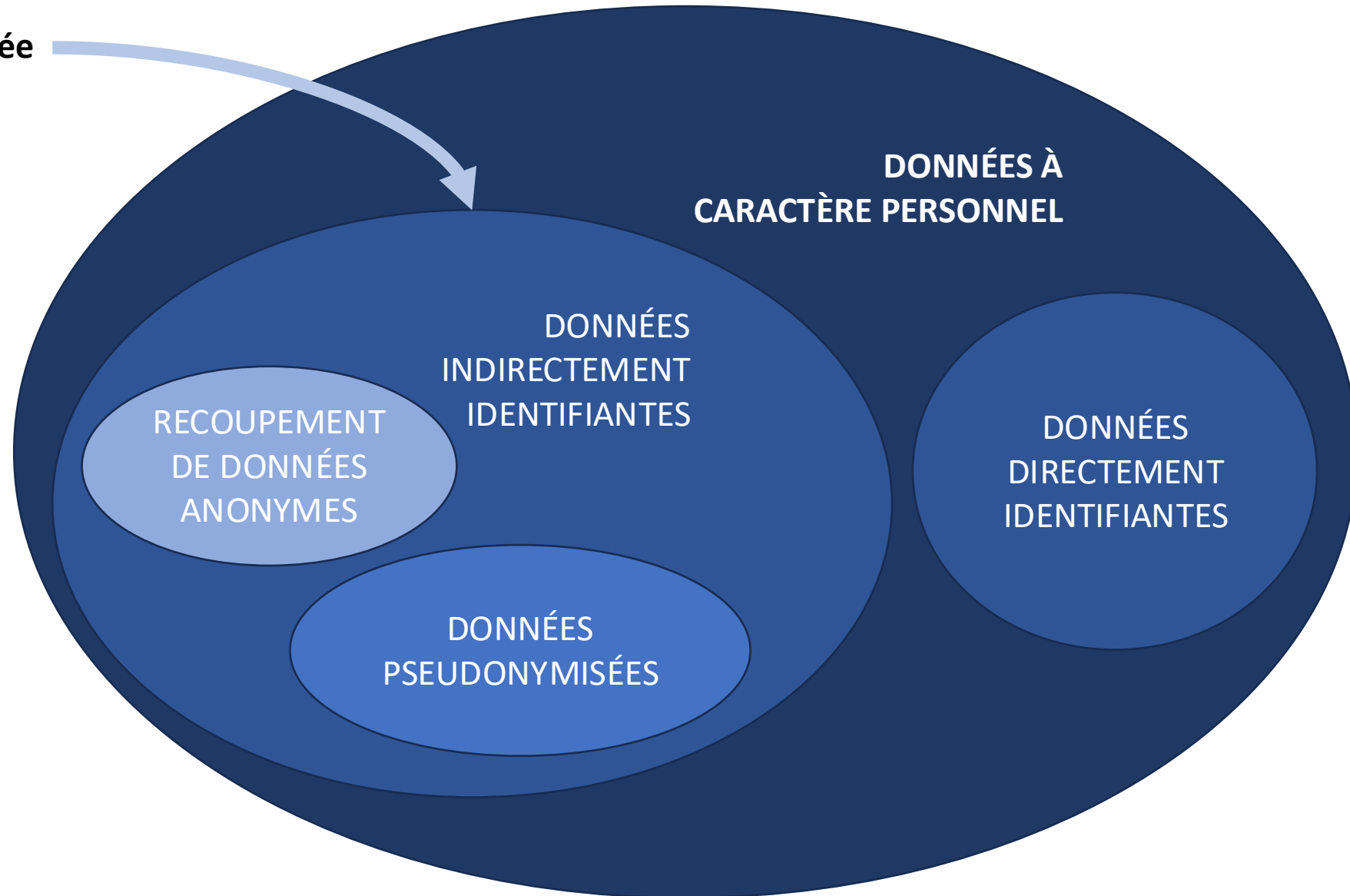
Pour résumer

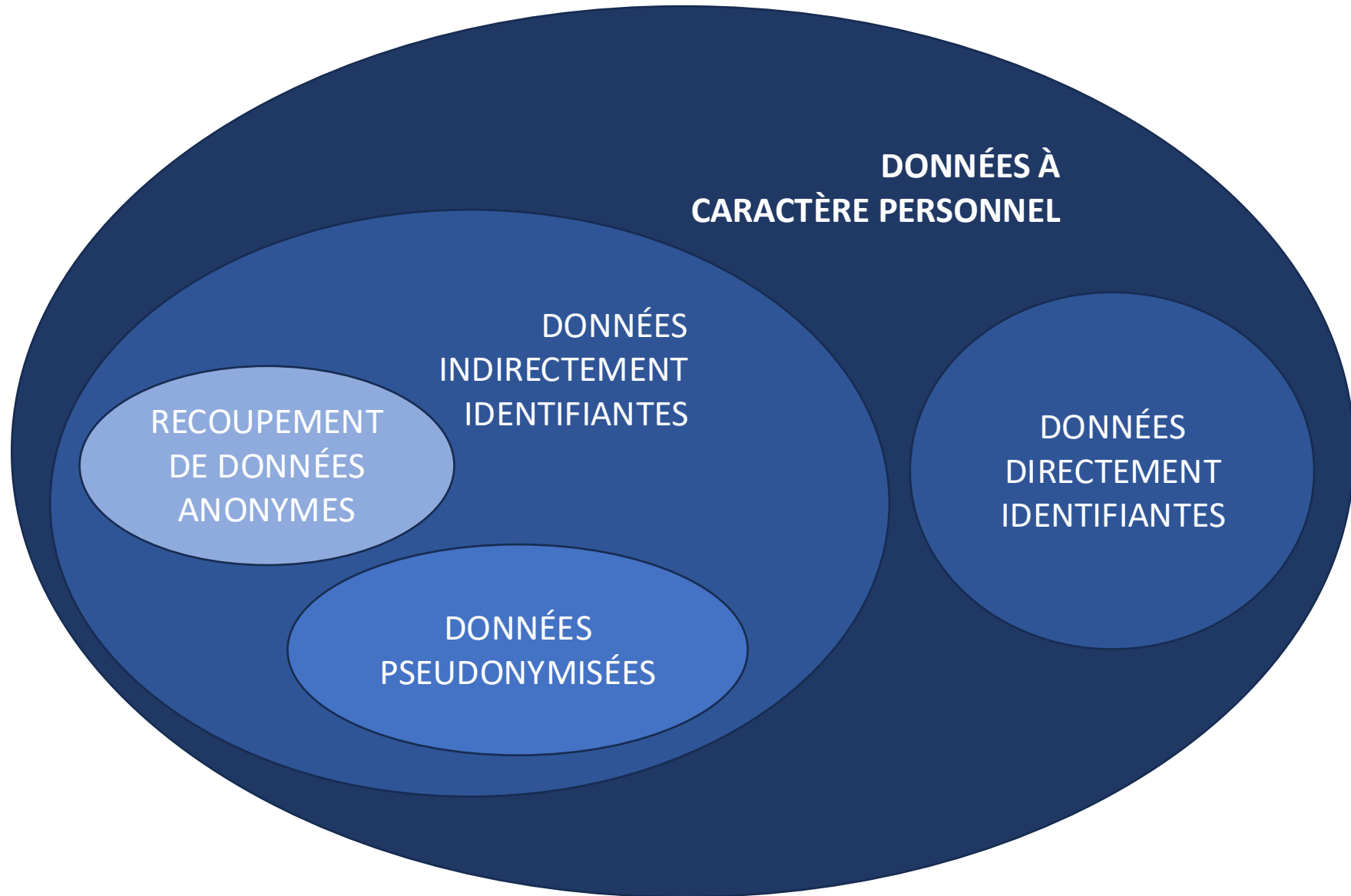
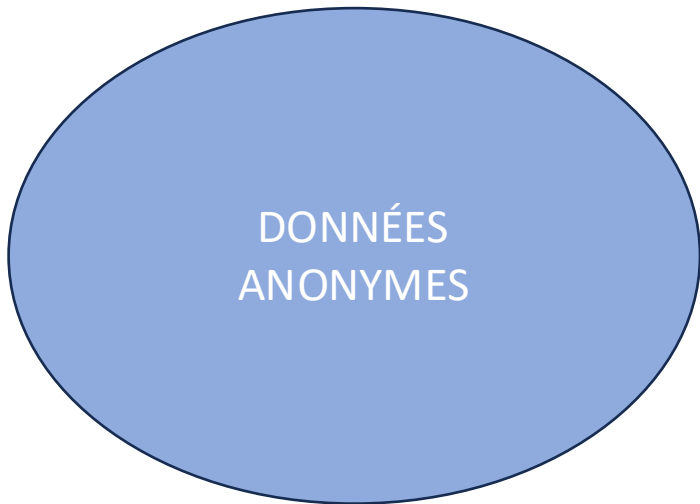
**DONNÉES À
CARACTÈRE PERSONNEL**



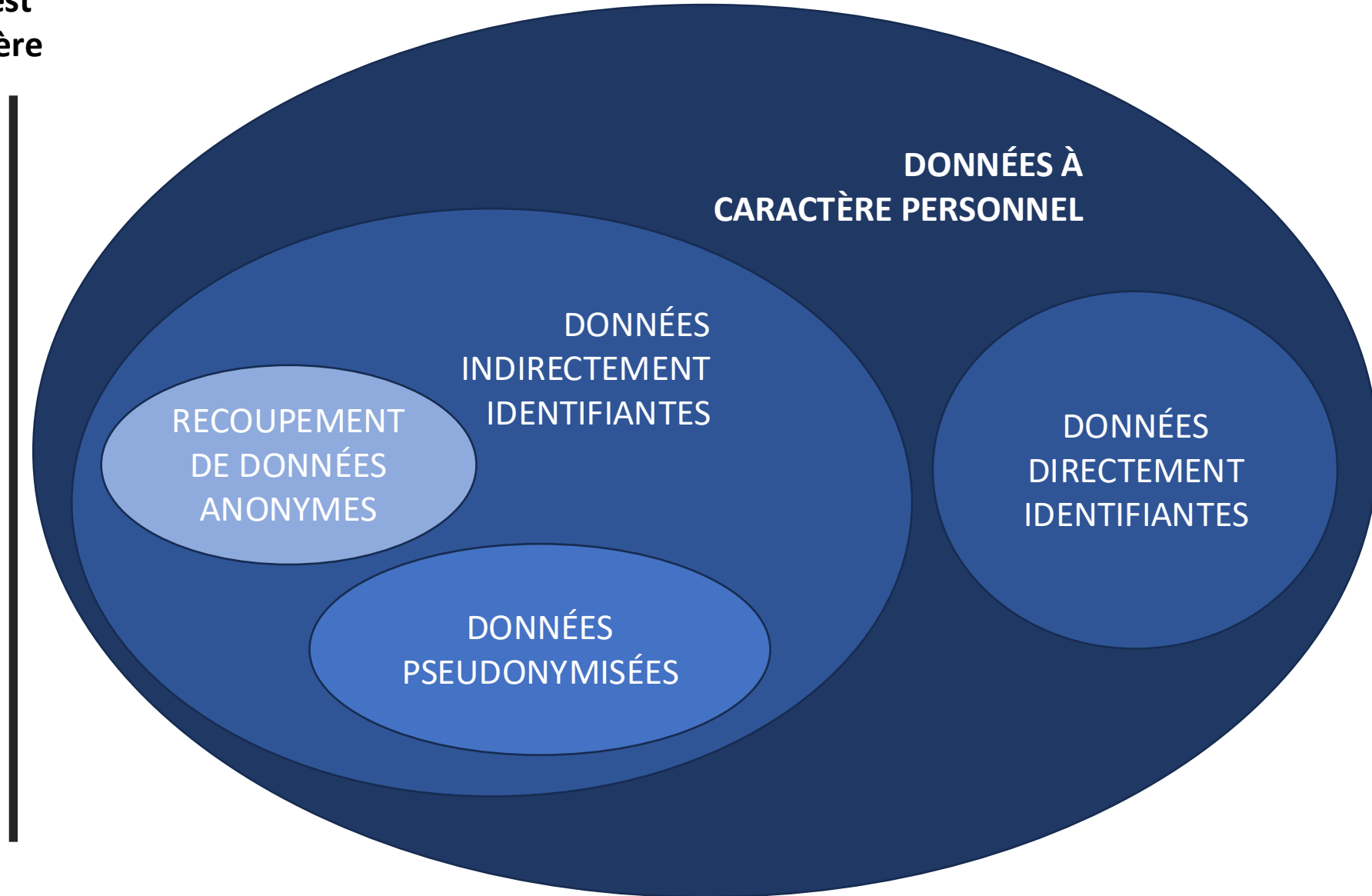
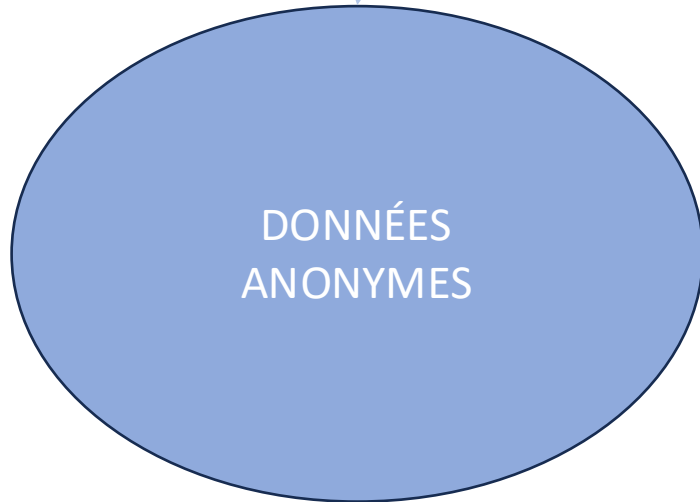


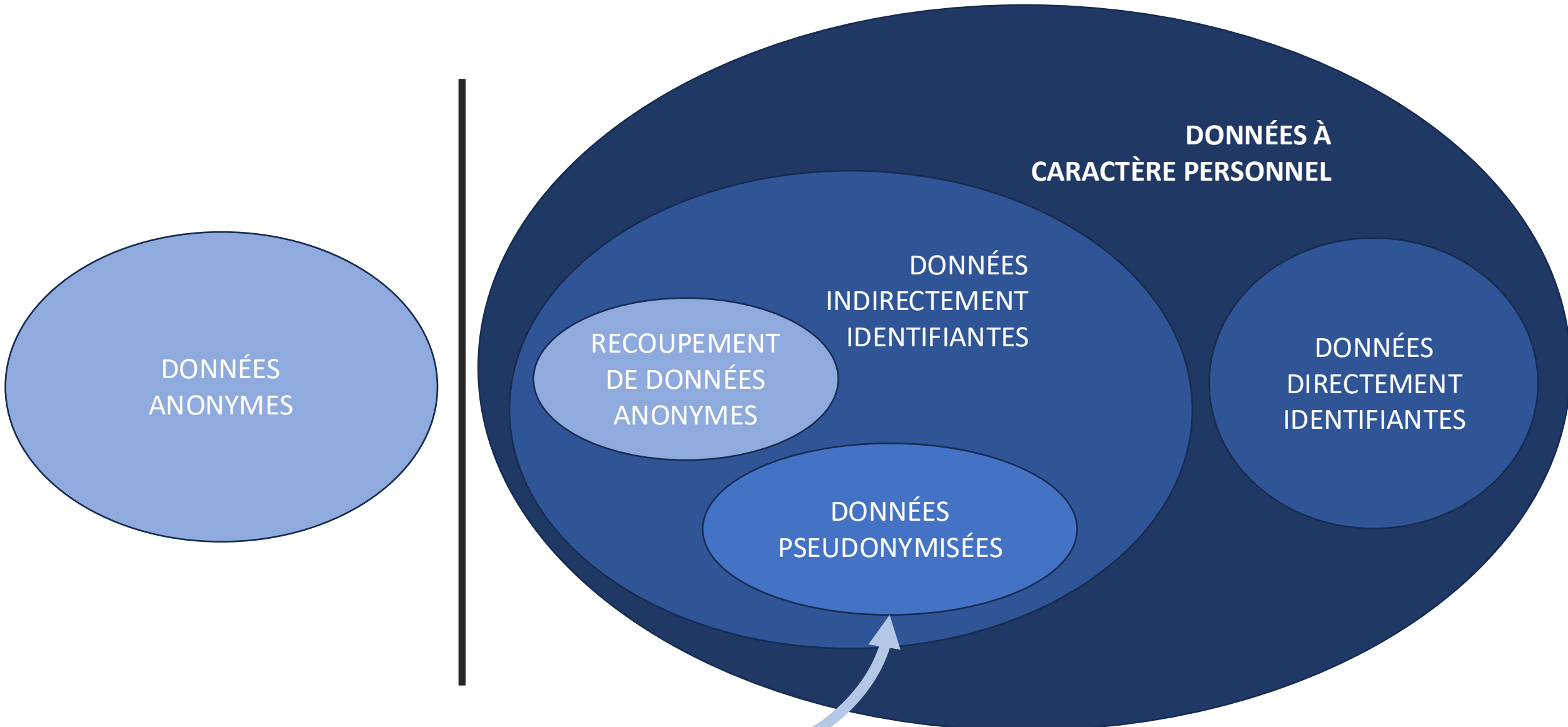
Une donnée non directement
identifiante peut être une donnée
à caractère personnel





Une donnée « anonyme » n'est pas/plus une donnée à caractère personnel





**Une donnée pseudonymisée
n'est pas une donnée « anonyme »**

TAKE HOME MESSAGES

Il existe différents types de données :

- Les données “à caractère personnel” (ou données personnelles) qui font référence à une personne
- Certaines de ces données sont dites “sensibles”, comme par exemple les données concernant la santé
- Il existe 3 catégories de données de santé (par nature, par combinaison et par destination)

Ces données peuvent avoir différents niveaux permettant l'identification de personnes

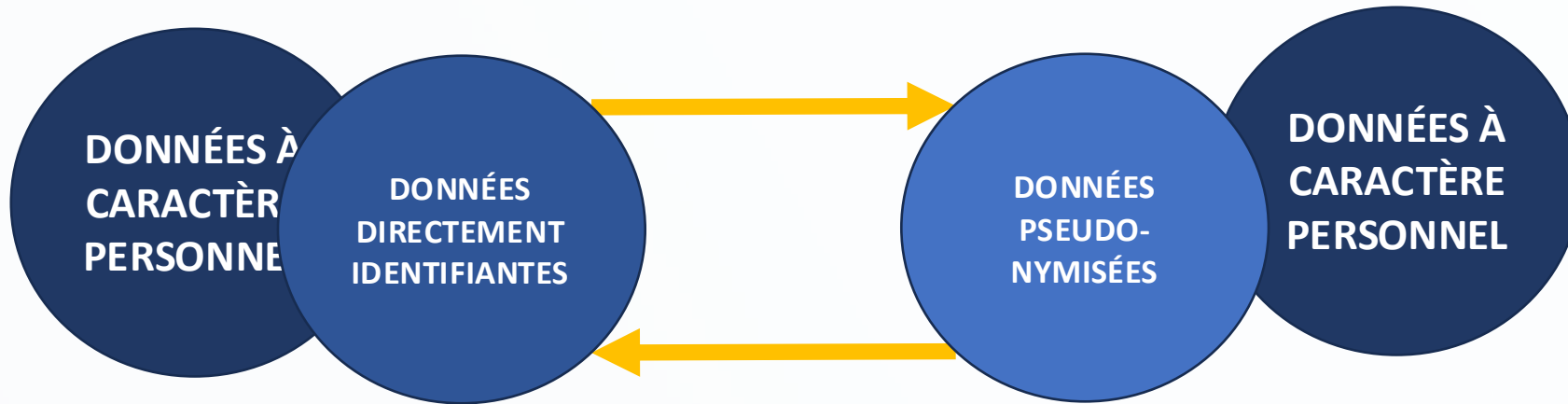
- directement identifiantes *ex. données nominatives*
- non directement identifiantes *ex. données pseudonymisées*
- non identifiantes *ex. données anonymes*



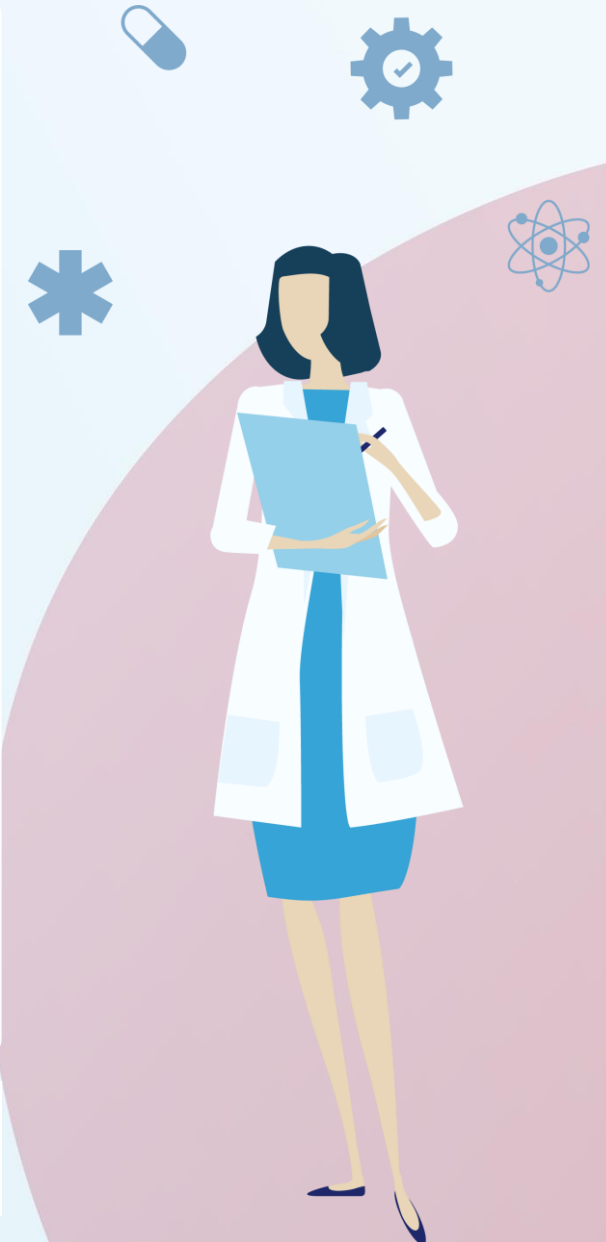
TAKE HOME MESSAGES

Pseudonymiser un jeu de données protège la vie privée (dé-identification) mais cela **ne le rend pas anonyme**.

Il permet le chaînage et l'appariement des données personnelles d'un individu.



Le processus inverse de re-identification est possible : la pseudonymisation est **reversible**.



POUR APPROFONDIR

Souces diverses pour la préparation de ce support :

La protection des données de santé, Véronique CABANES et Manon de FALLOIS,
service de la sante de la CNIL

https://esante.gouv.fr/sites/default/files/media_entity/documents/la-protection-des-donnees-de-sante.pdf

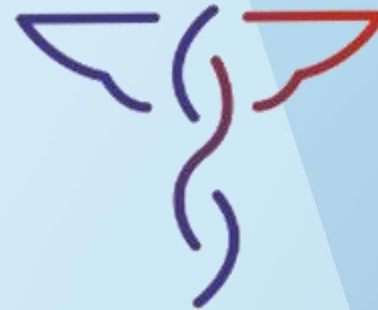
Introduction à la protection des données personnelles de santé, Julien Grosjean,
D2IM/Limics

https://www.cismef.org/cismef/wp/wp-content/uploads/2022/11/Introduction-RGPD_2022.pdf

Définitions du RGPD (ch.1, art.4) :

<https://eur-lex.europa.eu/legal-content/FR/TXT/?uri=celex%3A32016R0679>

+ références et liens en bas des slides



SN@SU
Santé Numérique
Sorbonne Université

*Ce travail a bénéficié d'une aide de l'État gérée par l'Agence Nationale de la Recherche au titre **de France 2030** portant la référence ANR-23-CMAS-0001*

