

1. Exercice

Fin Juin 2022, un groupe d'amis a décidé de faire un don du sang.

Voici une table regroupant les données personnelles que les médecins leur ont demandé avant leur prise de sang :



Nom	Date Naissance	Ville
Roux Emmanuelle	13/01/2006	Ardres
Masson Edeline	13/03/2007	Calais
Astier Cécile	01/03/2006	Calais
Boulot Céline	29/06/2006	Coulogne
Campagne David	16/05/2006	Ardres
Bernard Jean-Pierre	12/05/2006	Calais
Lefebvre Claire	21/02/2005	Calais
Mayeux Eric	13/06/2005	Coulogne
Garcia Michel	29/06/2006	Coulogne
Naïdji Antoine	13/04/2006	Marck
Van Helsen Baptiste	03/03/2005	Ardres

On appelle descripteur le nom associé aux données d'une colonne. Un descripteur est en plus un **identifiant** lorsqu'il possède une valeur unique, c'est-à-dire différente sur chaque ligne.

1. Quels sont les descripteurs utilisés ici ?

2. Certains de ces descripteurs sont-ils des identifiants pour ces données ? Si oui, lesquels ?

En fin de journée, l'infirmier a récapitulé les résultats des analyses dans un tableau sans mettre les noms pour les statistiques du laboratoire.

Age	Sexe	Code Postal	Groupe Sanguin
16	F	62137	A
16	G	62610	A
17	F	62100	A
16	F	62610	AB
16	G	62730	AB
17	G	62137	B
16	G	62137	B
17	G	62610	B
15	F	62100	B
16	G	62100	O
16	F	62100	O



Cours SNT	Thème : Données structurées Activité 4. Comment recouper des données ?	Date :
-----------	--	--------

3. Quels sont les descripteurs utilisés ici ?

4. Certains descripteurs sont-ils des identifiants ? Justifier.

5. Les données de cette table vous semblent-elles *anonymes* ?

L'ordinateur du secrétariat du laboratoire a été victime d'un piratage. Le pirate a récupéré les deux tables. En recoupant les données, il va être capable de *désanonymiser* les données médicales.

6. Quelles informations manquantes sont nécessaires pour mettre en relation les deux tables ? Va-t-il les trouver facilement ? Expliquer et les donner.

7. Retrouver le groupe sanguin de chaque patient en remplissant la table ci-dessous à gauche.
N.B : La table de droite contient les codes postaux de certaines villes du Pas-de-Calais.

Age	Sexe	Code Postal	Nom	Groupe Sanguin
16	F	62137	<input type="text"/>	A
16	G	62610	<input type="text"/>	A
17	F	62100	<input type="text"/>	A
16	F	62610	<input type="text"/>	AB
16	G	62730	<input type="text"/>	AB
17	G	62137	<input type="text"/>	B
16	G	62137	<input type="text"/>	B
17	G	62610	<input type="text"/>	B
15	F	62100	<input type="text"/>	B
16	G	62100	<input type="text"/>	O
16	F	62100	<input type="text"/>	O

Table des données désanonymisées

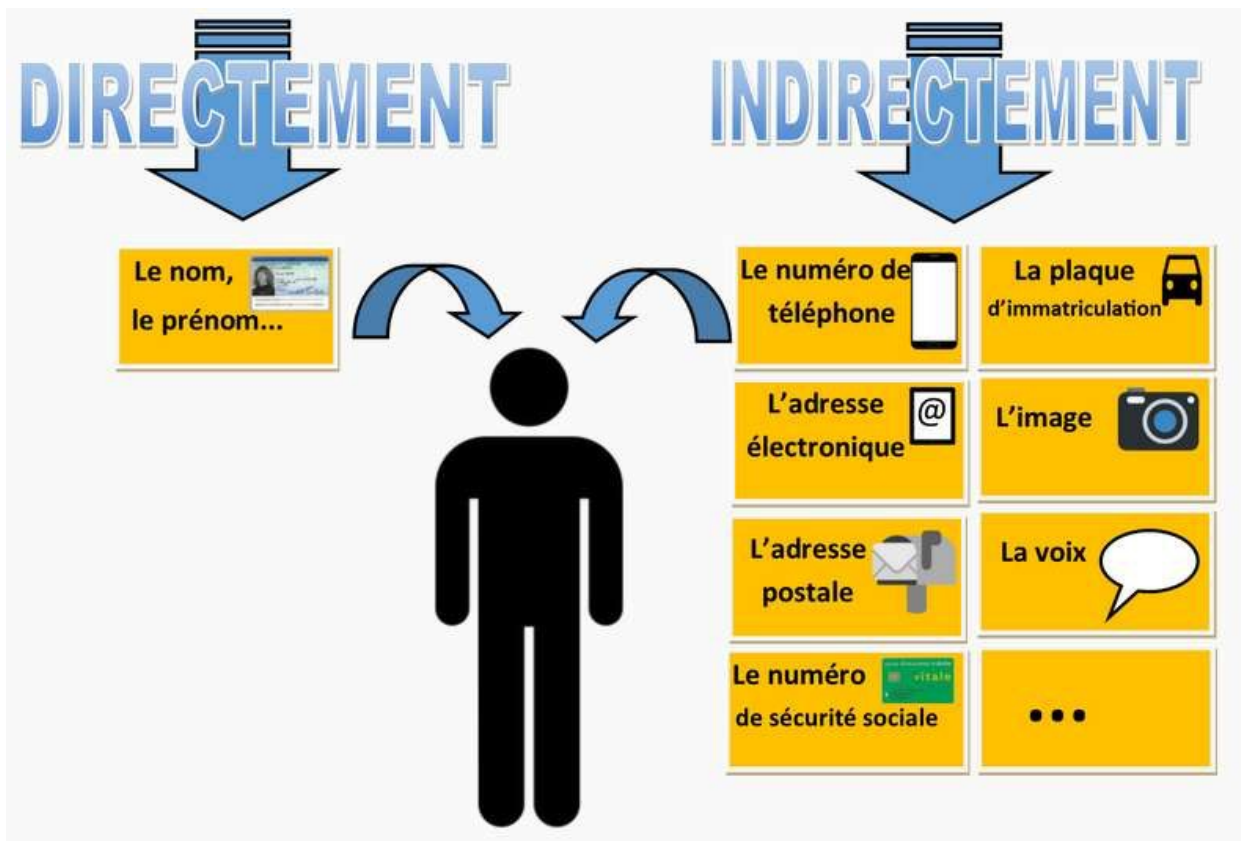
Code Postal	Ville
62000	Arras
62100	Calais
62137	Coulogne
62610	Ardres
62730	Marck

Table des codes postaux

2. Définition

Donnée à caractère personnel : Une donnée est dite « à caractère personnel » si elle se rapporte à une personne identifiable. Elle peut être de différente nature, textuelle ou non. Elle doit permettre d'identifier une personne directement, grâce à un identifiant, ou indirectement, par recoupement de plusieurs informations.

La [page du site de la CNIL](https://www.cnil.fr/fr/la-cnil) donne des exemples de données à caractère personnel.



Document 1. Qu'est-ce qu'une donnée personnelle ? Source : <https://blogacabdx.ac-bordeaux.fr/unum64/>

Le RGPD (Règlement Général de la Protection des Données) encadre le traitement des données personnelles sur le territoire de l'Union européenne et impose certaines obligations en matière de recueil de données, notamment :

- Recueillir l'**accord préalable** des clients
- **Inform**er les clients de leurs droits d'accès, de rectification, d'opposition et de suppression des informations collectées
- Veiller à la **sécurité** des systèmes d'information
- Assurer la **confidentialité** des données
- Indiquer une **durée de conservation** des données

Cours SNT	Thème : Données structurées Activité 4. Comment recouper des données ?	Date :
-----------	--	--------

Lorsqu'un organisme collecte des données personnelles, même si elles sont anonymisées, il est souvent possible, en croisant des données, de les désanonymiser. Une grande partie des individus d'une population est identifiable en ne connaissant que :

- leur date de naissance ou quelques dates précises (date d'entrée/sortie d'un hôpital, ...)
- leur sexe
- la zone géographique dans laquelle ils habitent

Lors de la création d'une base de données confidentielles et sensibles, comme celle des dossiers médicaux, il faut donc choisir avec un grand soin les descripteurs si l'on souhaite qu'il soit impossible, ou du moins très difficile, de désanonymiser ces données.

3. Synthèse

 **À Faire :** Compléter le texte à trou suivant

Une est « toute information se rapportant à une personne physique identifiée ou identifiable » de manière directe ou indirecte. Par exemple : , , .

Le est un dispositif de l'Union Européenne pour encadrer le traitement des données personnelles.

L'anonymisation des données est une technique consistant .

La consiste à retrouver une personne à partir de ses données personnelles anonymisées.

Selon [certaines études](#), par un système de recoupement et de filtre, il est possible de retrouver 83% des personnes vivant aux états-unis à partir de 3 descripteurs : le genre, la date de naissance et le code postal.