



# PROJET 4

Collecte des données dans le respect du RGPD



# Recommandations RGPD

5 règles de gestions à mettre en place pour être en conformité



## RGPD

Le règlement Général sur la Protection des Données (RGPD) est une législation Européenne entrée en vigueur le 25 mai 2018.

Il vise à renforcer et harmoniser la protection des données personnelles au sein de l'Union Européenne (UE).



# Recommandation RGPD n°1

5 règles de gestions à mettre en place pour être en conformité

## Recommandation 1

Supprimer et ne plus collecter les données de santé des clients. Sauf cas d'usages spécifiques, le traitement de ces données est interdit (art. 9 du RGPD).

Données concernées dans la base clients :

- groupe\_sanguin
- num\_ss



### Article 9 - Traitement portant sur des catégories particulières de données à caractère personnel

1. Le traitement des données à caractère personnel qui révèle l'origine raciale ou ethnique, les opinions politiques, les convictions religieuses ou philosophiques ou l'appartenance syndicale, ainsi que le traitement des données génétiques, des données biométriques aux fins d'identifier une personne physique de manière unique, des données concernant la santé ou des données concernant la vie sexuelle ou l'orientation sexuelle d'une personne physique sont interdits.



# Recommandation RGPD n°2

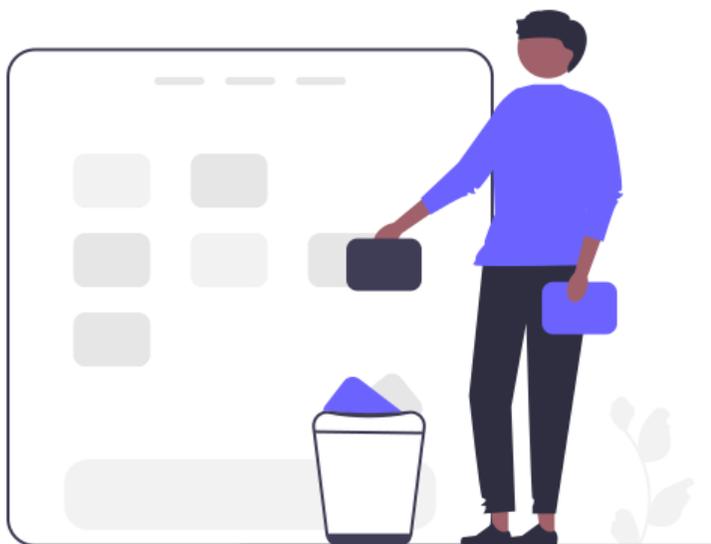
5 règles de gestions à mettre en place pour être en conformité

## Recommandation 2

Supprimer et ne plus collecter les données qui ne sont pas strictement nécessaires à atteindre la finalité du traitement des données (art. 5 du RGPD).

Données concernées dans la base de données :

- lat (latitude de la résidence du client)
- lon (longitude de la résidence du client)
- valeur\_residence\_prin (valeur résidence principale)



### Article 5 - Principes relatifs au traitement des données à caractère personnel

1. Les données à caractère personnel doivent être :

- a) traitées de manière licite, loyale et transparente au regard de la personne concernée (licéité, loyauté, transparence);
- b) collectées pour des finalités déterminées, explicites et légitimes, et ne pas être traitées ultérieurement d'une manière incompatible avec ces finalités; le traitement ultérieur à des fins archivistiques dans l'intérêt public, à des fins de recherche scientifique ou historique ou à des fins statistiques n'est pas considéré, conformément à l'article 89, paragraphe 1, comme incompatible avec les finalités initiales (limitation des finalités);
- c) adéquates, pertinentes et limitées à ce qui est nécessaire au regard des finalités pour lesquelles elles sont traitées (minimisation des données);



# Recommandation RGPD n°3

5 règles de gestions à mettre en place pour être en conformité

## Recommandation 3

Limiter la durée de conservation des données.

Il s'agit d'un principe général énoncé à l'article 5 du RGPD qui impose à chaque responsable de traitement de déterminer une durée de conservation des données personnelles cohérente et justifiée au regard de l'objectif de leur traitement.

Un organisme ne peut donc conserver des données personnelles de manière illimitée, sauf dans certains cas spécifiques et limités à ce qui est strictement nécessaire.

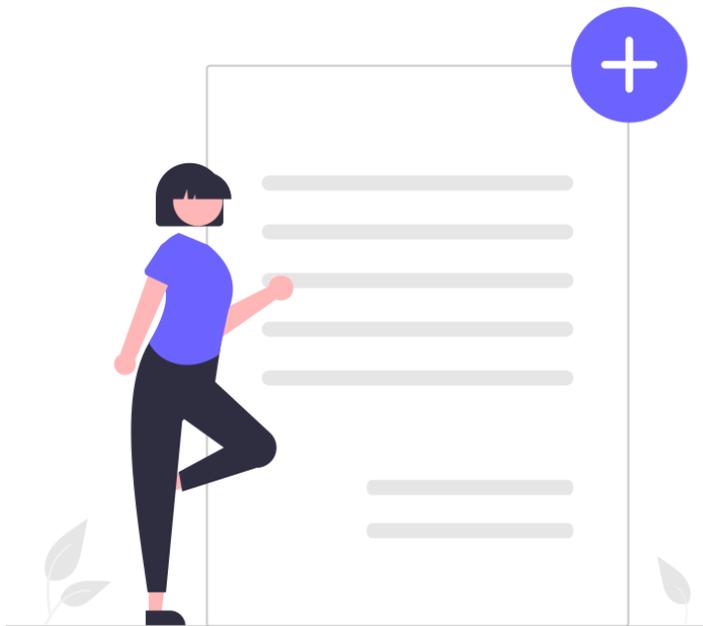
Source : [https://www.cnil.fr/sites/cnil/files/atoms/files/guide\\_durees\\_de\\_conservation.pdf](https://www.cnil.fr/sites/cnil/files/atoms/files/guide_durees_de_conservation.pdf)





# Recommandation RGPD n°4

5 règles de gestions à mettre en place pour être en conformité



## Recommandation 4

Documenter la conformité de l'entreprise en matière de conservation de données.

La documentation constitue un élément clé de conformité aux règles relatives à la protection des données personnelles : les actions entreprises, les analyses menées et les documents réalisés à chaque étape doivent être regroupés et centralisés pour la prouver.

Source : [https://www.cnil.fr/sites/cnil/files/atoms/files/guide\\_durees\\_de\\_conservation.pdf](https://www.cnil.fr/sites/cnil/files/atoms/files/guide_durees_de_conservation.pdf)



# Recommandation RGPD n°5

5 règles de gestions à mettre en place pour être en conformité

## Recommandation 5

Sécuriser les données et leur traitement (art 32 du RGPD).

S'assurer de la mise en place des mesures de sécurité techniques nécessaires à préserver l'intégrité, la disponibilité et la confidentialité des données.

Pour le traitement, compiler dans un document SI de référence « Politique de sécurité des systèmes d'informations (PSSI) », les procédures, contrôles et sensibilisations misent en place pour garantir la sécurité des données.



### Article 32 - Sécurité du traitement

1. Compte tenu de l'état des connaissances, des coûts de mise en œuvre et de la nature, de la portée, du contexte et des finalités du traitement ainsi que des risques, dont le degré de probabilité et de gravité varie, pour les droits et libertés des personnes physiques, le responsable du traitement et le sous-traitant mettent en œuvre les mesures techniques et organisationnelles appropriées afin de garantir un niveau de sécurité adapté au risque, y compris entre autres, selon les besoins:

- a) la pseudonymisation et le chiffrement des données à caractère personnel;
- b) des moyens permettant de garantir la confidentialité, l'intégrité, la disponibilité et la résilience constantes des systèmes et des services de traitement;
- c) des moyens permettant de rétablir la disponibilité des données à caractère personnel et l'accès à celles-ci dans des délais appropriés en cas d'incident physique ou technique;
- d) une procédure visant à tester, à analyser et à évaluer régulièrement l'efficacité des mesures techniques et organisationnelles pour assurer la sécurité du traitement.



# Requête SQL

Documentation des traitements effectués

```
select metier, sexe, date_naissance, id_client,  
enfant_conduite_accompagne, nombre_enfants, revenus,  
usage_vehicule, type_vehicule, est_rouge, points_perdus,  
age_vehicule, type_conduite, date_demande, etat_dossier, formule,  
tarif_devis, adresse  
from base_client  
where etat_dossier = 'complet' and date_demande like '2022%'
```

## Données écartées :

- employeur : inutile
- num\_ss : donnée sensible
- groupe\_sanguin : donnée médicale
- id\_site\_web : inutile
- nom : anonymisation par le retrait
- email : anonymisation par le retrait
- valeur\_residence\_prin : inutile
- formation : inutile
- lat : inutile
- lon : inutile



# Import des données CSV

Documentation des traitements effectués avec Power Query

## Choisir une source de données

Tous Fichier Services en ligne Autre

**Classeur Excel**  
Importez des données d'un classeur Microsoft Excel.

**Texte/CSV**  
Importez des données à partir d'un fichier texte ou CSV.

**XML**  
Importez des données d'un fichier XML.

**JSON**  
Importez des données à partir d'un fichier JSON.

**Liste SharePoint Online**  
Importez des données à partir d'une liste Microsoft SharePoint Online.

**OData**  
Importez des données à partir d'un flux OData.

**Table vide**  
Copiez et collez les données dans la table, ou entrez-les manuellement.

## Prévisualiser les données du fichier

Chemin d'accès au fichier : /Users/alexandre/SynologyDrive/Drive/Alex/documents/3 - formation en cours/Data Analyst/Open Classroom/Data Analyst - Cours/04 - Collectez des donn...

Origine du fichier: 65001: Unicode (UTF-8) Délimiteur: Virgule Détection du type de données: Selon les 200 premières lignes

metier	sexe	date_naissa...	id_cli...	enfant_conduite_accomp...	nombre_enfa...	reven...	usage_vehic...	type_vehic...
animateur du patrimoine	F	25/09/1999	871024631	0	0	125301	Commercial	Sports Car
préparateur en pharmacie	F	28/01/1932	7945239	1	1	50815	Private	Minivan
électricien installateur installatrice	F	15/09/1978	713692930	0	2	21204	Private	Pickup
chargé de veille législative et réglementaire	F	17/02/2021	5888885...	0	0	0	Private	z_SUV
conseiller pénitentiaire d'insertion et de probation	M	14/06/1956	69165559	0	0	null	Commercial	Panel Truck
staffeur-ornemaniste	F	11/04/1992	573751437	0	1	79606	Private	z_SUV
technicien électrotechnicien électrotechnicienne	F	18/03/1941	294816298	2	2	130540	Commercial	Minivan
gestionnaire de patrimoine	M	06/03/1956	686791659	0	0	51278	Private	z_SUV
officier de gendarmerie	M	18/02/1956	8660933...	0	0	115044	Private	Minivan
peintre en bâtiment	F	07/03/1974	401200262	0	0	null	Commercial	Panel Truck
intégrateur web	F	13/02/1967	926073051	1	2	24060	Private	Minivan
coordonnateur d'études cliniques	M	14/05/2017	45830738	0	4	20822	Private	z_SUV

Précédent

Annuler

Transformer les données

Charger



# Transformation des données #1

Documentation des traitements effectués avec Power Query

## # anonymisation

**Colonne personnalisée** ?

Ajoutez une colonne calculée en fonction d'autres colonnes ou valeurs.

Nouveau nom de colonne \*

 Type de données 

Formule de colonne personnalisée \* ⓘ

```
= Number.RoundDown(Duration.Days((DateTime.Date(DateTime.LocalNow()))-[date_naissance]) / 365)
```

Colonne(s) disponible(s)

- metier
- sexe
- date\_naissance
- id\_client
- enfant\_conduite\_accompagne
- nombre\_enfants
- revenu

[En savoir plus sur les formules Power Query](#)

**date\_naissance** remplacée par age

=> Insertion d'une colonne personnalisée incluant une formule permettant de calculer le nombre de jours séparant la date de naissance du jour actuel, valeur ensuite / 365 et arrondie à l'entier inférieur :  
« `Number.RoundDown(Duration.Days((DateTime.Date(DateTime.LocalNow()))-[date_naissance]) / 365)` »



# Transformation des données #2

Documentation des traitements effectués avec Power Query

## # anonymisation

### Colonne personnalisée ?

Ajoutez une colonne calculée en fonction d'autres colonnes ou valeurs.

Nouveau nom de colonne \*

Type de données

Nombre entier

Formule de colonne personnalisée \* ⓘ

= Number.RandomBetween(1000000, 10000000)

Colonne(s) disponible(s)

- metier
- sexe
- annee\_naissance
- id\_client
- enfant(s)\_en\_conduite\_acc...
- a\_des\_enfants
- revenu

Insérer une colonne

Annuler OK

[En savoir plus sur les formules Power Query](#)

**id\_client** celui-ci aurait pu être écarté dès l'extraction de la base (ma préconisation). Si toutefois nous le gardons (ce qui est le cas ici)

=> *Création d'une colonne personnalisée nommée « id\_client\_random » et utilisation de la formule « Number.RandomBetween(1000000, 10000000) » permettant de générer un chiffre aléatoire compris entre les 2 chiffres mis en parenthèse. Suppression de la colonne d'origine contenant les id.*



# Transformation des données #3

Documentation des traitements effectués avec Power Query

## # anonymisation

**enfant\_conduite\_accompagnee** et **nombre\_enfants** nombre remplacé par les valeurs Oui ou Non.

### Ajouter une colonne conditionnelle ?

Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne \*

enfant(s)\_en\_conduite\_accom...

	Nom de la colonne *	Opérateur *	Valeur ⓘ		Sortie ⓘ
Si	1 <sup>2</sup> 3 enfant_condui...	est égal à	0	Alors	Non
Autrement si	1 <sup>2</sup> 3 enfant_condui...	est supérieur à	0	Alors	Oui

Ajouter une clause

Else ⓘ



NC

=> *Insertion d'une colonne conditionnelle*

Annuler

OK



# Transformation des données #4

Documentation des traitements effectués avec Power Query

## # anonymisation

**revenus** remplacé par tranche\_de\_revenus  
([0-20k€], [20k€-50k€], [50k€-100k€],  
[100k€-200k€], [+200k€])

=> *Insertion d'une colonne conditionnelle*

### Ajouter une colonne conditionnelle ?

Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne \*

tranche\_de\_revenus

	Nom de la colonne *	Opérateur *	Valeur <span>?</span>	Alors	Sortie <span>?</span>
Si	1 <sup>2</sup> 3 revenus	est inférieur ou égal à	20000		[0-20k€]
Autrement si	1 <sup>2</sup> 3 revenus	est inférieur ou égal à	50000		[20k€-50k€]
Autrement si	1 <sup>2</sup> 3 revenus	est inférieur ou égal à	100000		[50k€-100k€]
Autrement si	1 <sup>2</sup> 3 revenus	est inférieur ou égal à	200000		[100k€-200k€]
Autrement si	1 <sup>2</sup> 3 revenus	est supérieur à	200000		[+ de 200k€]

Ajouter une clause

Else ?

NC

Annuler

OK



# Transformation des données #5

Documentation des traitements effectués avec Power Query

## # anonymisation

En savoir plus

### Diviser la colonne ?

Spécifiez le délimiteur utilisé pour fractionner la colonne de texte.

De base  Avancé

Séparateur \*

Personnalisé. ▾

/

Annuler OK

**date\_demande** précise (horodatée)  
remplacée par mois\_demande

=> *division de la colonne par délimiteur « / » puis par délimiteur « espace » pour écartier l'heure. Fusion des colonnes mois et année et suppression des colonnes inutiles.*



# Transformation des données #6

Documentation des traitements effectués avec Power Query

## # anonymisation

**Texte après le délimiteur**

Entrez le délimiteur qui marque le début des éléments à extraire.

De base  Avancé

Délimiteur \*

Rechercher le délimiteur

À partir de la fin de l'entrée

Nombre de délimiteurs à ignorer ⓘ

1

Annuler OK

**adresse** remplacée par localisation (moins précise) ne contenant aucune information sur la rue et le numéro

=> *extraction de texte après délimiteur, en partant de la fin et en ignorant un espace pour ne garder que la fin de l'adresse contenant la ville (et éventuellement le CP).*



# Transformation des données #7

Documentation des traitements effectués avec Power Query

## # données tarifaires

**tarif\_devis** remplacé par tranche\_tarif\_devis  
 ([-250€], [250€ à 300€], [300€ à 350€], [350€ à 400€], [400€ à 500€], [+500€],)

### Ajouter une colonne conditionnelle

Ajoutez une colonne conditionnelle calculée en fonction des autres colonnes ou valeurs.

Nouveau nom de colonne \*

tranche\_tarif\_devis

	Nom de la colonne *	Opérateur *	Valeur 	Alors	Sortie 
Si	1 <sup>2</sup> 3 tarif_devis.1	est inférieur à	250	Alors	[- de 250€]
Autrement si	1 <sup>2</sup> 3 tarif_devis.1	est inférieur ou égal à	300	Alors	[de 250 à 300€]
Autrement si	1 <sup>2</sup> 3 tarif_devis.1	est inférieur ou égal à	350	Alors	[de 300 à 350€]
Autrement si	1 <sup>2</sup> 3 tarif_devis.1	est inférieur ou égal à	400	Alors	[de 350 à 400€]
Autrement si	1 <sup>2</sup> 3 tarif_devis.1	est inférieur ou égal à	500	Alors	[de 400 à 500€]
Autrement si	1 <sup>2</sup> 3 tarif_devis.1	est supérieur à	500	Alors	[+ de 500€]

Ajouter une clause

Else 

NC

Annuler

OK

=> *division de la colonne par délimiteur (.)  
 pour arrondir et permettre la reconnaissance  
 en chiffre puis insertion colonne conditionnelle*

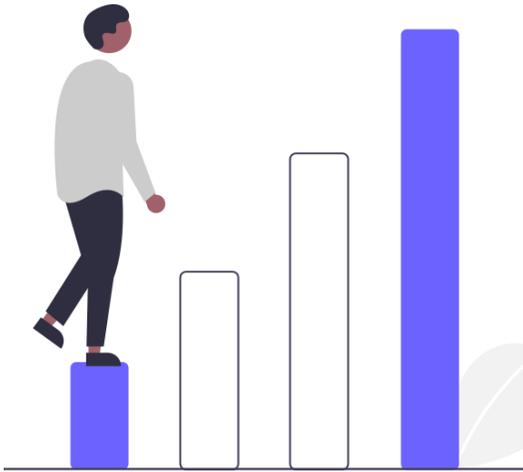


# Transformation des données #8

Documentation des traitements effectués avec Power Query

## # qualité des données

J'ai profité du traitement des données dans powerQuery pour uniformiser les données en sortie. Ci-dessous le détail des traitements effectués.



- la colonne **est\_rouge** utilise des réponses en anglais (yes/no). Je la remplace par une colonne `est_rouge_opt` avec des réponses en français (oui/non)  
*=> Insertion d'une colonne conditionnelle*
- la colonne **age\_vehicule** comporte un chiffre décimal (.0). Je la simplifie en retirant la décimale. Nouvelle colonne renommée en `age_du_vehicule`.  
*=> division de la colonne avec délimiteur (.) et suppression de la partie inutile*



**END**

Merci pour votre écoute 😊