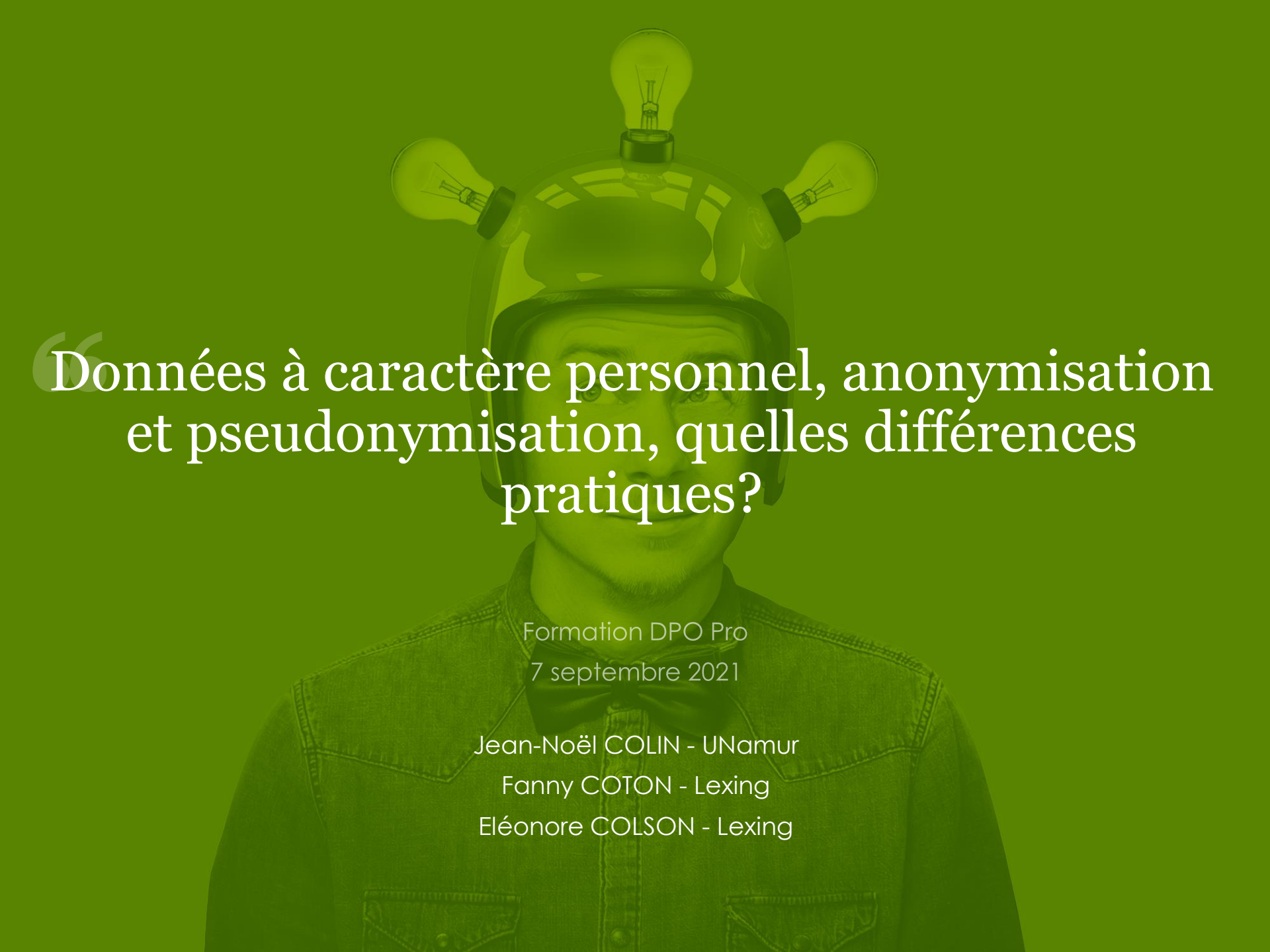


Données à caractère personnel, anonymisation et pseudonymisation, quelles différences pratiques?

Formation DPO Pro

7 septembre 2021

Jean-Noël COLIN - UNamur

Fanny COTON - Lexing

Eléonore COLSON - Lexing

Notre fil conducteur

- La **société pharmaceutique** Harold Goldfarb est basée en **Australie**.
- Elle prend contact avec l'hôpital Hilltop (BE) pour **tester** un nouveau médicament destiné à soigner les patients atteints de la Covid 19.
- Le Professeur Andromeda est désigné comme « **investigateur principal** » de cet essai clinique.
- Le projet de contrat d'étude soumis par la société prévoit que la société a accès aux données de l'étude, en ce compris les **données des patients**, sujets de l'essai.

Programme

1. **Données à caractère non personnel, donnée anonyme ou donnée pseudonyme, comment les distinguer ?**

2. **Comment mettre en place l'anonymisation des données à caractère personnel ?**

3. **Traitement de données à caractère non personnel : quelles sont les règles à respecter ?**

4. **Synthèse et enjeux actuels**

Quel est votre profil ?

Informaticien

Juriste

1.
Données à caractère non personnel, données anonymes, données pseudonymes, … comment les distinguer ?

Donnée à caractère personnel

Toute information se rapportant à une **personne physique** identifiée **ou identifiable**

Ex : nom, numéro d'identification, données de localisation, plaque d'immatriculation, identifiant en ligne, adresse IP, résultats médicaux, e-mails, affiliations, …

→ Lorsqu'une personne peut être **isolée** dans des ensembles de données
(Pas nécessaire de connaître son nom)

→ RGPD applicable

Données à caractère **non** personnel

Données à caractère non personnel

= un concept **négatif**

Données autres que les données à caractère personnel **telles que définies par le RGPD**

Ex : *« des sets de données agrégées et anonymisées utilisées pour des analyses Big Data »*

Règlement 2018/1807 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union

Notre fil conducteur

Dans le cadre de l'essai clinique, les données suivantes des patients sont traitées :

- Date de naissance ;
- Initiales ;
- Code postal ;
- Numéro d'identification assigné dans le cadre de l'étude ;
- Poids ;
- Taille ;
- Historique des tests Covid19 ;
- Réaction du patient au traitement ;
- Particularités physiques ;
- Comorbidités ;
- Traitements médicaux éventuels ;
- Séquelles de la Covid19

Notre fil conducteur

- Lors des discussions à propos du contrat, Harold Goldfarb réitère son souhait de **recevoir toutes les données « en clair ».**
- Le Professeur Andromeda est interpelé par cette exigence. Il demande **l'avis du DPO** de l'hôpital Hilltop.

Principe de minimisation

- **Art. 5.1.e du RGPD** « les données à caractère personnel doivent être conservées sous une forme permettant l'identification des personnes concernées pendant une durée n'excédant pas celle nécessaire au regard des finalités pour lesquelles elles sont traitées ».
  - *Accountability*
- **Art. 198 de la loi du 30 juillet 2018** : Obligation dans le cadre de la recherche scientifique, historique ou à des fins statistiques
  - *« Lors d'un traitement de données à des fins de recherche scientifique ou historique ou à des fins statistiques fondé sur une collecte de données auprès de la personne concernée,* ***le responsable du traitement anonymise ou pseudonymise les données après leur collecte****. »*
  - *Accountability* ***renforcée***

Anonymisation/Pseudonymisation : **Définitions**

Procédé à l'issue duquel les données à caractère personnel traitées …

- ne peuvent plus être attribuées à une personne concernée précise **sans avoir recours à des informations supplémentaires**

  ➔ Informations conservées séparément et de manière sûre, en ayant recours à des mesures techniques et organisationnelles (telles que le chiffrement)

PSEUDONYMISATION

- ne peuvent plus être attribuées à une personne donnée, **même en utilisant des données supplémentaires.**

ANONYMISATION

Lignes directrices relatives au règlement concernant un cadre applicable au libre flux des données à caractère non personnel dans l'Union européenne

Notre fil conducteur

- Le jeune stagiaire graphiste de l'hôpital, Monsieur Neutron, est chargé de l'anonymisation des données des patients.

- Sa technique est simple :
  - copier les données des sujets de l'étude issues de la DB patients de l'hôpital dans un fichier excel,
  - puis retirer du set de données les numéros d'identification (et ne conserver que les autres données de poids, taille, etc.)

Attention à la **combinaison** de bases de données

- *« (…) Un ensemble de données* ***considérées comme anonymes*** *peut être* ***combiné avec un autre ensemble de données*** *de telle façon qu'un ou plusieurs individus deviennent* ***identifiables****. »* (Groupe 29, Avis 05/2014 sur les techniques d'anonymisation)

- Justice de Paix du canton de Forest, 9 juin 2020
  - Données de pointage des abonnements MOBIB
  - Selon la STIB, le RGPD ne s'applique pas car la DB des données de pointage ne contient que les numéros de carte MOBIB
  - Juge : **le RGPD s'applique** car **corrélation possible** de la DB des données de pointage avec la DB clients, même si séparée

Attention aux possibilités d'inférence

- 2006 : Le scandale AOL : données de **650.000** utilisateurs

The New York Times

A Face Is Exposed for AOL Searcher No. 4417749

By MICHAEL BARBARO and TOM ZELLER Jr. AUG. 9, 2006

Buried in a list of 20 million Web search queries collected by AOL and recently released on the Internet is user No. 4417749. The number was assigned by the company to protect the searcher's anonymity, but it was not much of a shield.

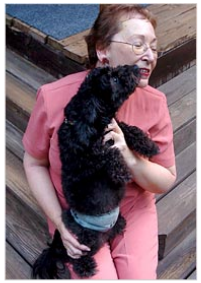
No. 4417749 conducted hundreds of searches over a three-month period on topics ranging from "numb fingers" to "60 single men" to "dog that urinates on everything."

And search by search, click by click, the identity of AOL user No. 4417749 became easier to discern. There are queries for "landscapers in Lilburn, Ga," several people with the last name Arnold and "homes sold in shadow lake subdivision gwinnett county georgia."

It did not take much investigating to follow that data trail to Thelma Arnold, a 62-year-old widow who lives in Lilburn, Ga., frequently researches her friends' medical ailments and loves her three dogs. "Those are my searches," she said, after a reporter read part of the list to her.

AOL removed the search data from its site over the weekend and apologized for its release, saying it was an unauthorized move by a team that had hoped it would benefit academic researchers.

In the privacy of her four-bedroom home, Ms. Arnold searched for the answers to scores of life's questions, big and small. How could she buy "school supplies for Iraq children"? What is the "safest place to live"? What is "the best season to visit Italy"?



Thelma Arnold's identity was betrayed by AOL records of her Web searches, like ones for her dog, Dudley, who clearly has a problem.
Erik S. Lesser for The New York Times

Her searches are a catalog of intentions, curiosity, anxieties and quotidian questions. There was the day in May, for example, when she typed in "termites," then "tea for good health" then "mature living," all within a few hours.

Her queries mirror millions of those captured in AOL's database, which reveal the concerns of expectant mothers, cancer patients, college students and music lovers. User No. 2178 searches for "foods to avoid when breast feeding." No. 3482401 seeks guidance on "calorie counting." No. 3483689 searches for the songs "Time After Time" and "Wind Beneath My Wings."

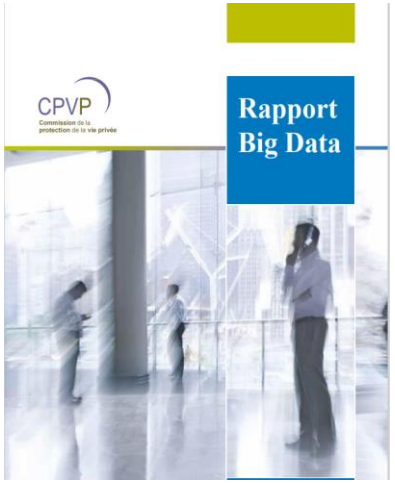
There are also many thousands of sexual queries, along with searches about "child porno" and "how to kill oneself by natural gas" that raise questions about what legal authorities can and should do with such information.

But while these searches can tell the casual observer — or the sociologist or the marketer — much about the person who typed them, they can also prove highly misleading.

Quels sont les textes / recommandations ?

Avis 05/2014 sur les techniques d'anonymisation (WP 216)

- Pas de normes prescriptives dans la législation
- **Trois critères** essentiels pour déterminer la robustesse de chaque technique :



- **Individualisation**
- **Corrélation**
- **Inférence**

- suppression d'éléments directement identifiants ne suffit pas en soi
- mesures supplémentaires en fonction du contexte et des finalités du traitement

Position de l'Autorité de Protection des Données sur l'anonymisation des images de vidéosurveillance

- APD, 24/2021
  - Les images sont anonymes, l'identification des passants est rendue impossible
  - Grâce à anonymisation « quasi immédiate » des données :
    - ✓ *Privacy by design/by default*
    - ✓ Minimisation des données

Afbeelding 1. Beelden in lage resolutie weergegeven door de firmware (fase 1)

Afbeelding 2. Weggebruikers vervangen door "blobs" (fase 2)

2.
Comment mettre en place l'anonymisation des données à caractère personnel ?

Sondage

Actuellement, procédez-vous :

 De la pseudonymisation

 De l'anonymisation

Big Data

- Une explosion de données
  - d'immenses quantités de données sont collectées,
  - à partir de sources très différentes (incl. réseaux sociaux, terminaux mobiles (Coronalert 🤔), IoT),
  - dans des formats très variés, structurés ou non (image, texte, son, vidéo)
  - fournies par l'utilisateur ou capturées ou inférées.
  - renforcée par le phénomène open data
- De nouveaux outils et méthodes
  - stockage
  - traitement
  - analyse

Un peu de vocabulaire

- [Dataset] jeu de données, ensemble de données, d'enregistrements
- [Attribut] information présente dans le dataset, caractérisant les enregistrements
- [Identifiant direct] attribut qui identifie un individu (empreinte digitale, EID)
- [Identifiant indirect ou quasi-identifiant] attribut qui n'identifie pas directement un individu, sauf s'il est combiné à d'autres informations (données sociales, géolocalisation…)
- [Attribut sensible] attribut dont on ne veut pas révéler la valeur
- [Non-identifiant] attribut ne permettant pas l'identification, ni directe, ni indirecte
- [Anonymisation] « 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. » (source https://www.cnil.fr)
- [Pseudonymisation] « traitement de données personnelles réalisé de manière à ce qu'on ne puisse plus attribuer les données relatives à une personne physique sans information supplémentaire. » (source https://www.cnil.fr)

Big Data

- Difficultés
  - Une fois publiée, il est impossible de 'rattraper' une donnée
  - Une information rendue publique peut être composée, croisée, corrélée avec d'autres informations publiques
  - Le concept de 'donnée anonyme' est relatif
- Quels risques pour la vie privée?
  - [Individualisation] possibilité d'isoler une partie ou la totalité des enregistrements identifiant un individu dans le dataset
  - [Corrélation] possibilité de relier entre eux au moins deux enregistrements se rapportant au même individu ou groupe d'individus
  - [Inférence] possibilité de déduire avec une forte probabilité la valeur d'un attribut à partir des valeurs d'autres attributs

Anonymiser les données

- Comment?
  - Transformer les données originales de manière à interdire toute identification (directe ou indirecte) de la personne concernée
  - Selon un processus irréversible
  - Tout en préservant l'utilité de la donnée
- Anonymisation statique vs dynamique
  - [Statique] les données sont anonymisées avant d'être publiées
  - [Dynamique] les données retournées sont anonymisées à la volée, en fonction de la requête d'accès
- Deux grandes approches
  - Appauvrir les données
  - Dégrader les données

Appauvrir les données

- **Suppression**
  - de record ou d'attribut
  - possibilité de créer un attribut de substitution (ex. durée d'emploi remplace les dates de début et de fin de contrat)
  - impact possible sur l'utilisabilité du dataset, en particulier d'un point de vue statistique

- **Masquage**
  - remplacement de caractères dans les valeurs d'un attribut (ex. j*****c@gmail.com)
  - applicable uniquement pour les attributs textuels
  - préserver ou non la longueur?

Dégrader les données

- **Généralisation**
  - remplacer les valeurs d'un attribut par des plages de valeurs (ex. remplacer 29 par 20-30) ou des catégories (faible, moyen, élevé)
  - attention au nombre d'enregistrements par classe (outlier)
  - impact possible sur l'utilisabilité du dataset car perte d'information

- **Pseudonymisation**
  - remplacement d'un attribut identifiant par une valeur opaque, dérivée ou non de la valeur originale
  - faut-il préserver le lien entre enregistrements?
  - pseudonyme aléatoire, chiffré, hash. . . (🤔 réellement à sens unique?)
    - Ex. Jean-Noël Colin ⇒ e115e5dcb395a7b644e40212e680ccd13ba018f2d461ccdb989d069af19430e8

Dégrader les données

- **Permutation**
  - Permuter les valeurs d'un attribut entre enregistrements
  - Les caractéristiques statistiques sont préservées
  - Quid d'une éventuelle relation entre attributs pour un même enregistrement? ok pour analyse intra-attribut
  - Impact possible sur l'utilisabilité du dataset

- **Bruitage**
  - Modification de la valeur d'un attribut en y ajoutant un bruit, une distorsion (ex. arrondi, ajout d'une valeur aléatoire)
  - Impact possible sur l'utilisabilité du dataset, en particulier d'un point de vue statistique

Dégrader les données

- **Ajout de données synthétiques**
  - Créer des données artificielles pour compléter le dataset
  - Les caractéristiques statistiques sont-elles préservées?
  - Quid d'une éventuelle relation entre attributs pour un même enregistrement? ok pour analyse intra-attribut
  - Impact possible sur l'utilisabilité du dataset

- **Agrégation**
  - Regroupement d'enregistrements en valeurs agrégées
  - Utile uniquement si pas besoin des valeurs individuelles

Anonymiser les données

- Approche générale

  - Déterminer le modèle de publication (privé/public)
  - Identifier la nature des données à anonymiser (identifiant, quasi identifiant, sensible…)
  - Identifier les besoins de l'utilisateur des données
  - Anonymiser les données en fonction des besoins
  - **Évaluer le risque de ré-identification (encore et encore…)**
  - Ajuster les mesures complémentaires (légales et organisationnelles)

Rendre les données anonymes

N° identifiant	Code postal	Date de naissance	Genre	Co-morbidité
55549476724	5346	30/05/1999	M	obésité
56704162551	1240	18/09/1990	M	obésité
68419995689	1022	13/05/1986	M	obésité
69450014461	5345	03/03/1984	M	hypertension
56832915413	1224	15/06/1982	M	maladie hématologique
43203076277	1241	08/09/1979	M	diabète
21305437922	3000	24/06/1974	F	obésité
28050471787	1000	09/09/1967	F	hypertension
55630480496	5396	06/01/1954	F	diabète
42564265436	1050	01/08/1947	F	hypertension

Données brutes

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	5346	21	M	obésité
2	1240	30	M	obésité
3	1022	34	M	obésité
4	5345	37	M	hypertension
5	1224	38	M	maladie hématologique
6	1241	41	M	diabète
7	3000	46	F	obésité
8	1000	53	F	hypertension
9	5396	67	F	diabète
10	1050	73	F	hypertension

Suppression de l'identifiant, ajout d'un pseudonyme, généralisation date de naissance

Rendre les données anonymes

- *k*-anonymat
  - anonymiser les données de manière à ce qu'un enregistrement ne puisse être distingué de min. *k-1* autres
  - créer classes d'équivalence comportant min. *k* enregistrements
  - le risque de ré-identification devient donc *1/k*
  - supprimer éventuellement les enregistrements isolés, ou appliquer une anonymisation plus forte
  - plus *k* est grand, plus la perte d'information est élevée

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	**	20-34	M	obésité
2	**	20-34	M	obésité
3	**	20-34	M	obésité
4	**	35-49	M	hypertension
5	**	35-49	M	maladie hématologique
6	**	35-49	M	diabète
7	**	35-49	F	obésité
8	**	50-99	F	hypertension
9	**	50-99	F	diabète
10	**	50-99	F	hypertension

Première tentative de *k*-anonymat

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	**	20-34	M	obésité
2	**	20-34	M	obésité
3	**	20-34	M	obésité
4	**	35-49	M	hypertension
5	**	35-49	M	maladie hématologique
6	**	35-49	M	diabète
7	**	35-49	F	obésité
8	**	50-99	F	hypertension
9	**	50-99	F	diabète
10	**	50-99	F	hypertension

Première tentative de *k*-anonymat

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	**	20-34	M	obésité
2	**	20-34	M	obésité
3	**	20-34	M	obésité
4	**	35-45	M	hypertension
5	**	35-45	M	maladie hématologique
6	**	35-45	M	diabète
7	**	46-99	F	obésité
8	**	46-99	F	hypertension
9	**	46-99	F	diabète
10	**	46-99	F	hypertension

Données 3-anonymes

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	**	20-34	M	obésité
2	**	20-34	M	obésité
3	**	20-34	M	obésité
4	**	35-45	M	hypertension
5	**	35-45	M	maladie hématologique
6	**	35-45	M	diabète
7	**	46-99	F	obésité
8	**	46-99	F	hypertension
9	**	46-99	F	diabète
10	**	46-99	F	hypertension

Données 3-anonymes

Rendre les données anonymes

- *l*-diversité
  - si au sein d'une classe d'équivalence, les valeurs d'un attribut ne sont pas suffisamment diversifiées, il est possible d'inférer la valeur de cet attribut pour les membres de cette classe
  - ex: dans un fichier patients *k*-anonymisé, dans la classe des patients vivant dans le namurois (classe d'équivalence), tous les enregistrements ont une pathologie pulmonaire. On peut en déduire que si Bob figure dans le fichier et habite Namur, il est atteint de cette maladie
  - la *l*-diversité requiert d'avoir au sein de chaque classe d'équivalence au moins l valeurs différentes pour les attributs sensibles

Rendre les données anonymes

Pseudonyme	Code postal	Age	Genre	Co-morbidité
1	**	20-45	M	obésité
2	**	20-45	M	obésité
3	**	20-45	M	obésité
4	**	20-45	M	hypertension
5	**	20-45	M	maladie hématologique
6	**	20-45	M	diabète
7	**	46-99	F	obésité
8	**	46-99	F	hypertension
9	**	46-99	F	diabète
10	**	46-99	F	hypertension

Données 3-anonymisées et 3-diversifiées

Rendre les données anonymes

- *t*-proximité
  - la *t*-proximité requiert que la distribution des valeurs d'un attribut sensible dans une classe soit sensiblement pareille à la distribution dans l'ensemble du dataset

Des outils pour *assister* la démarche

- Ateliers logiciels
  - ARX Data Anonymization Tool
  - Amnesia
  - µ-argus
  - Aircloak
  - IRI FieldShield
  - …
- Référentiels et normes
  - ENISA: Techniques et meilleures pratiques de pseudonymisation
  - ISO 20889:2017 Terminologie et classification des techniques de dé-identification de données pour la protection de la vie privée
    - Techniques de chiffrement préservant l'utilité des données: déterministe, préservant l'ordre, préservant le format, homomorphique, cherchable…
    - Anatomisation
    - Echantillonnage

Notre fil conducteur

Risque de réidentification ?

Quelques cas de ré-identification

- Sweeney, 2002
  - l'étude a mis en évidence qu'aux US, le triplet **<code postal, date de naissance, sexe>** permet d'identifier 87% de la population
  - ceci a permis de croiser une base de données médicales pseudonymisée avec une liste d'électeurs
  - Latanya Sweeney. 2000. Simple Demographics Often Identify People Uniquely. *Carnegie Mellon University, Data Privacy*. Retrieved from http://dataprivacylab.org/projects/identifiability/

Quelques cas de ré-identification

- Netflix, 2006
  - en 2006, Netflix a rendu publique une liste de 100 millions d'évaluations de films soumises par 500.000 utilisateurs entre 1999 et 2005, à des fins d'amélioration de son algorithme de recommandation
  - la liste reprenait (pseudonyme utilisateur, film, évaluation, date)
  - par croisement avec les évaluations disponibles sur le site IMDb, il a été démontré qu'il était possible de ré-identifier certains utilisateurs
  - ce qui implique dans certains cas la divulgation possible quant aux opinions politiques, religieuses ou autres centres d'intérêt personnels
  - Arvind Narayanan and Vitaly Shmatikov (2006). "How To Break Anonymity of the Netflix Prize Dataset". In: CoRR abs/cs/0610105. arXiv: cs/0610105. url: http://arxiv.org/abs/cs/0610105

Quelques cas de ré-identification

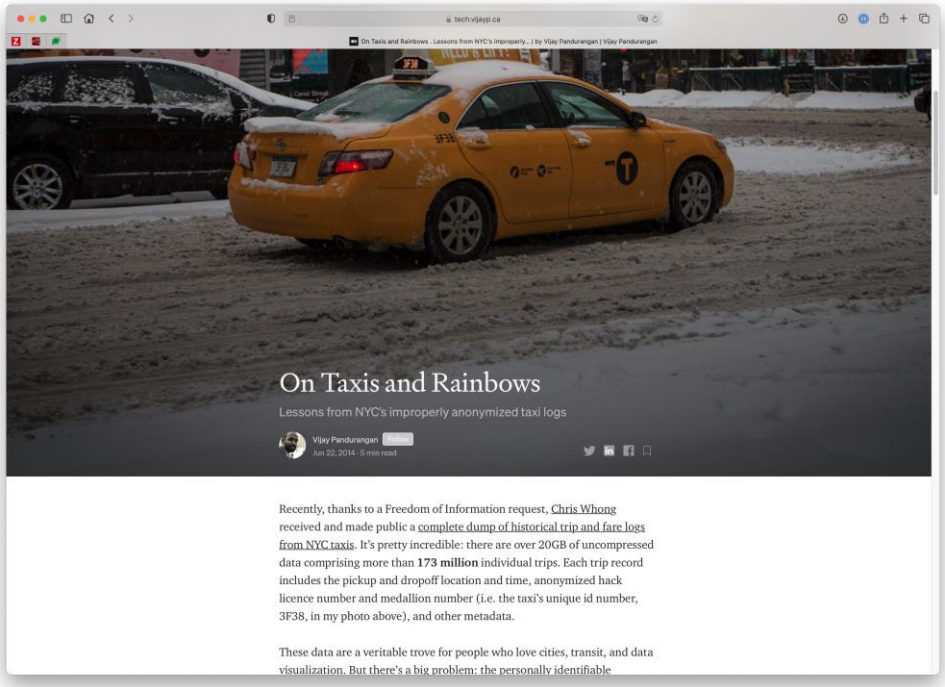
- Données de localisation
  - Des chercheurs ont montré que dans un fichier reprenant 15 mois de collecte de coordonnées spatio-temporelles d'1.5 million de personnes sur un rayon de 100km, il était possible d'isoler 95% de la population avec seulement 4 points et qu'avec seulement deux points, il était possible d'isoler plus de 50% des individus
  - Y.-A. de Montjoye et al. (Mar. 2013). “Unique in the Crowd: The privacy bounds of human mobility”. In: Scientific Reports 3, 1376, p. 1376. doi: 10.1038/srep01376

Quelques cas de ré-identification

- Réseaux sociaux
  - il a été montré que le graphe des relations d'un individu dans un réseau social pouvait servir d'identifiant
  - A. Narayanan and V. Shmatikov (2009). “De-anonymizing Social Networks”. In: 2009 30th IEEE Symposium on Security and Privacy, pp. 173–187. doi: 10.1109/SP.2009.22

Quelques cas de ré-identification

- Mauvais usage d'une fonction d'anonymisation (hash)



Source: https://tech.vijayp.ca/of-taxis-and-rainbows-f6bc289679a1

Conclusion

The Information That Is Needed to Identify You: 33 Bits

By WSJ Staff

Aug 4, 2010 12:20 am ET

With more than 6.6 billion people in the world, it's easy to feel safely anonymous.

Many data collectors assure consumers that they don't collect or store personally identifiable information — things like full names, Social Security numbers or credit-card numbers. But researchers say it's often possible to identify people even without that information.

The problem, says Paul Ohm, an associate professor at the University of Colorado Law School, is that there's a lot of data available about people. When clever statisticians tap those sources, Mr. Ohm says, "you should never bet against re-identification."

Publié dans le Wall Street Journal

$\log_2(6.600.000.000) = 32.62$
$\log_2(7.800.000.000) = 32.86$
$\log_2(11.200.000.000) = 33.38$

Tout serait alors **toujours identifiable ?**

RGPD Considérant 26 :

« En tenant compte des:

- Moyens **raisonnablement** susceptibles d'être utilisés
- Par le responsable du traitement **ou par toute autre personne**
- Pour identifier la personne **directement ou indirectement** »

→ Pas nécessaire que la donnée permette, À ELLE SEULE, d'identifier la personne concernée

→ TOUJOURS UN RISQUE RÉSIDUEL

→ Anonymisation impossible ?

Anonymisation impossible ?

- 83% des Américains peuvent être ré-identifiés à partir de leur genre, de leur date de naissance et de leur code postal.
- Les chercheurs (UCLouvain et Imperial College de Londres) sont partis d'un ensemble de données anonymisées et ont développé un algorithme de *machine learning* permettant d'identifier quels critères peuvent rendre une personne unique dans un groupe donné. (

  « The observatory of anonymity » https://cpg.doc.ic.ac.uk/observatory/

Rocher, L., Hendrickx, J.M. & de Montjoye, YA. Estimating the success of re-identifications in incomplete datasets using generative models. *Nat Commun* 10, 3069 (2019). https://doi.org/10.1038/s41467-019-10933-3

Notre fil conducteur

- L'un des sujets de l'étude, guéri de la Covid19 grâce au traitement, présente une séquelle de la Covid19 assez inhabituelle : ses yeux sont passés du brun au bleu.
- Les données de l'étude sont scrupuleusement anonymisées. Cette séquelle particulière y est mentionnée mais ne peut plus être attribuée à une personne physique.

Limiter le risque de manière **raisonnable**

G29 (CEPD) : Avis 05/2014 sur les Techniques d'anonymisation (WP216)

→ **Admet les limites de la technique:**

" Dans beaucoup de situations, un ensemble de données anonymisées peut encore présenter **un risque résiduel** pour les personnes concernées.

…

Aucune des techniques décrites dans le présent document ne satisfait de façon certaine aux critères d'une anonymisation efficace "

→ **Mais N'EXCLUT PAS la possibilité d'anonymiser**

→ **Conseille de :**

Évaluer les risques

Réévaluer régulièrement les risques

Limiter le risque de manière **raisonnable**

APD : Recommandation 03/2020 sur la destruction des données

- → Admet les limites de la technique
- → Pas suffisant comme mesure de destruction

- → **Mais conseille de faire tester le risque de réidentification par du personnel indépendant de celui qui a procédé à** l'anonymisation
  → garanties contractuelles à convenir

Moyens **raisonnables**

CJUE : arrêt Breyer, 19 octobre 2016 // avis AG

- Interprétation stricte de la notion de « moyens susceptibles d'être raisonnablement mis en œuvre »
- donner un sens au mot **« raisonnablement »**

IDENTIFIABLE SI :

- **moyens légaux**
- **possibilité d'une action judiciaire**

NON IDENTIFIABLE SI :

- Interdit par la loi (piratage)
- Irréalisable en pratique (effort démesuré en termes de temps, de main-d'œuvre et de coût)
- Possibilité négligeable ou purement hypothétique

3.
Traitement de données à caractère non personnel :

Notre fil conducteur

- Goldfarb décide d'étudier de **nouveaux paramètres** susceptibles d'affecter l'efficacité du traitement :
  - La durée quotidienne d'ensoleillement
  - La pollution atmosphérique

Règlement 2018/1807 établissant un cadre applicable au libre flux des données à caractère non personnel dans l'Union

Données à caractère non personnel =
un concept **negatif**

Données autres que les données à caractère personnel **telles que définies par le RGPD**

Régime

- Exigences de **localisation** des données (Art. 4) → 

- **Disponibilité** des données

  (Art. 5) Les autorités compétentes peuvent demander ou obtenir l'accès aux données pour l'accomplissement de leurs fonctions officielles, conformément au droit de l'Union ou au droit national

- **Portage** des données

  (Art. 6) La Commission encourage et facilite l'élaboration de codes de conduite par autorégulation au niveau de l'Union (ci-après dénommés «codes de conduite»), afin de contribuer à une économie des données compétitive, fondée sur les principes de transparence et d'interopérabilité

Quid en cas de traitement “mixte” de données ?

Données à caractère personnel + Données à caractère non personnel

- Si les données ne sont pas inextricablement liées :
  - Règlement 2018/1807 s’applique aux données non personnelles
  - RGPD s’applique aux données personnelles

- Si les données sont inextricablement liées :
  - *« Lorsque les données à caractère personnel et les données à caractère non personnel d'un ensemble sont inextricablement liées, le présent règlement est sans préjudice de l'application du règlement (UE) 2016/679. »*

    ➔ Le RGPD s’applique pour le tout

Notre fil conducteur

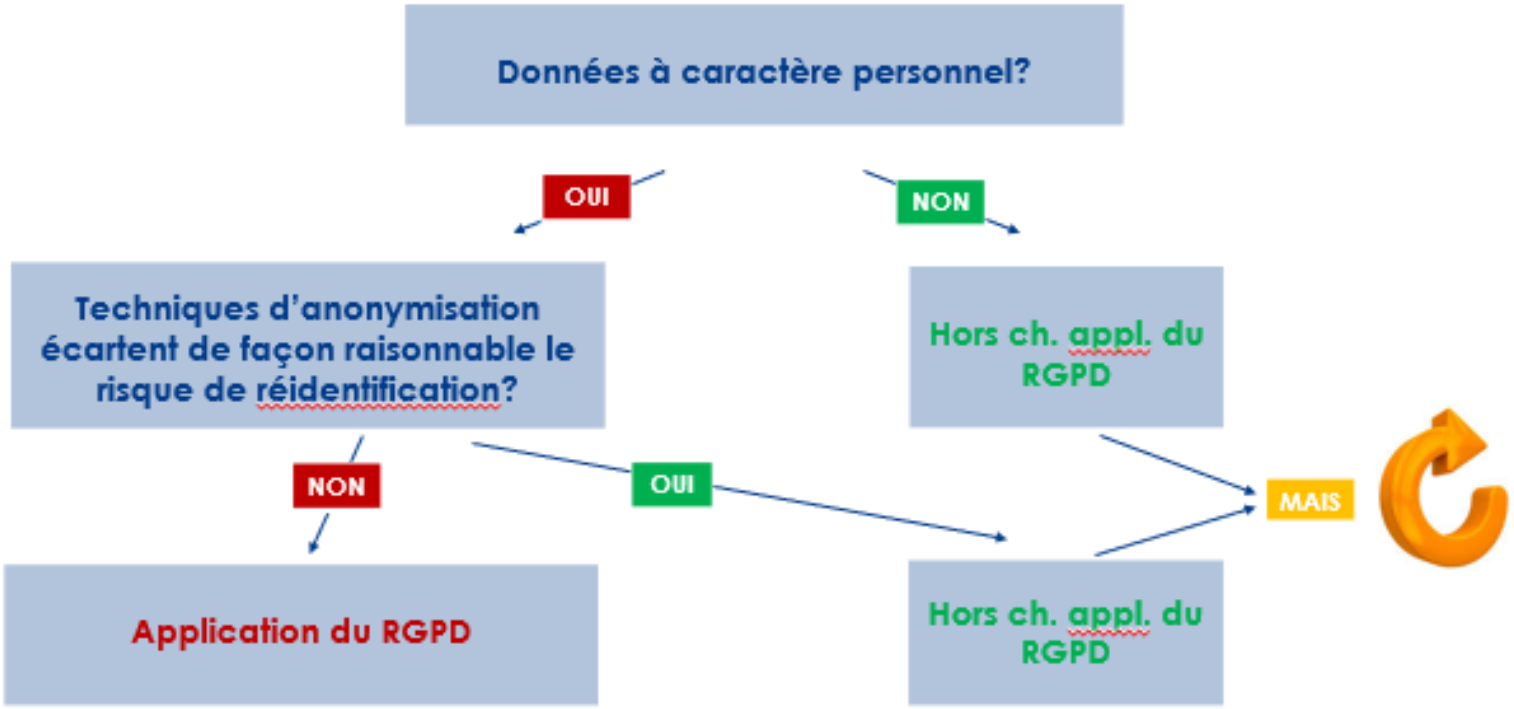
- La durée quotidienne d'ensoleillement
- La pollution atmosphérique

4.
Synthèse et enjeux actuels :

Synthèse

Données non personnelles		Données pseudonymisées	Données anonymisées
Depuis l'origine	Données initialement à caractère personnel		Données initialement à caractère personnel
Irréversible		**Réversible**	**Irréversible**
RGPD pas applicable		**RGPD applicable**	**RGPD pas applicable**

Schéma récapitulatif

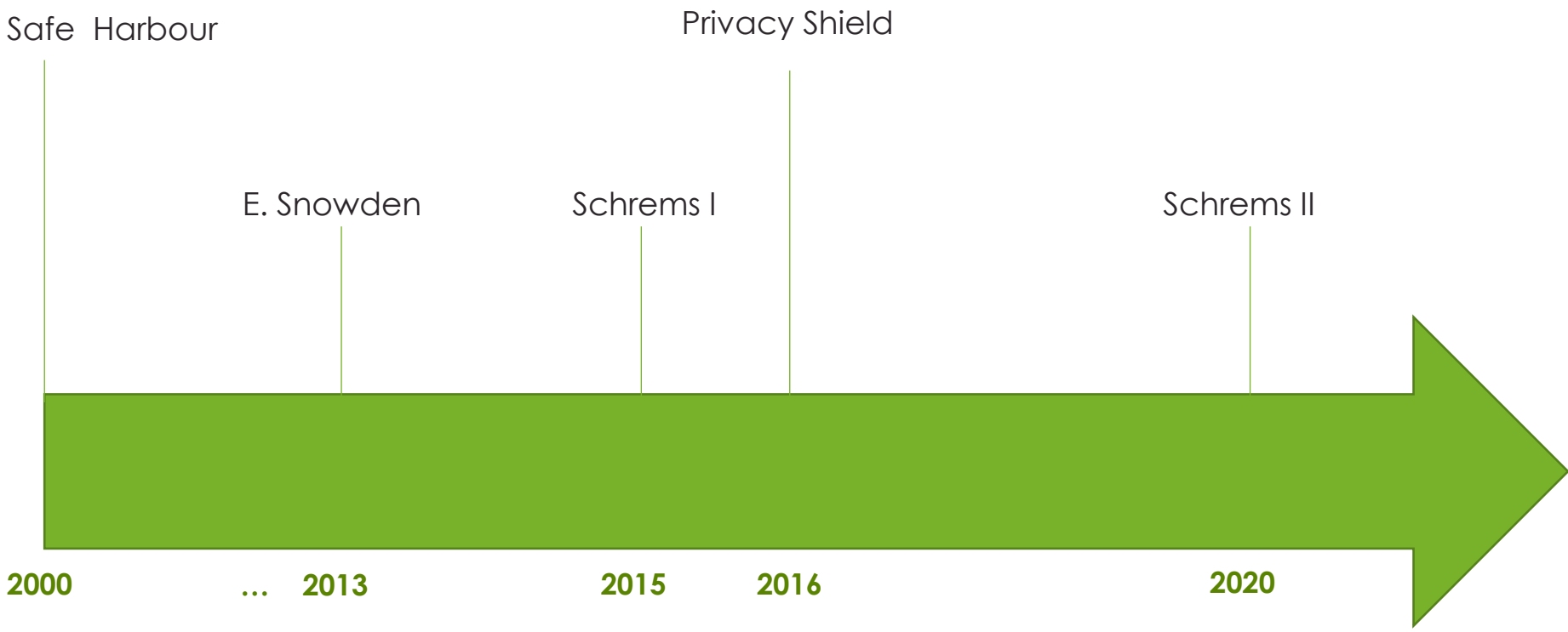


Quel est l'intérêt de pseudonymiser alors ?

- Sécurité (Groupe 29 Avis 05/2014 sur les Techniques d'anonymisation)
- Limitation du risque (et des hypothèses de notification) en cas de violation de données
- Data room
- Obligation dans le cadre de la recherche scientifique ou historique ou à des fins statistiques (Loi du 30 juillet 2018)

  + présomption de compatibilité d'un traitement ultérieur
- Transfert de données hors UE ?

Schrems I et Schrems II



Enseignements de Schrems II

1. Aucune décision d'adéquation pour le transfert de données vers les Etats Unis

2. Les CCT sont valides mais ne suffisent pas à elles seules

Il faut :

- **Analyser la législation du pays de destination**
- **Prendre des mesures supplémentaires :**

Mesures techniques supplémentaires

Pseudonymisation

- une analyse approfondie des données démontre qu'elles **ne peuvent être recoupées avec aucune information que les autorités publiques du pays destinataire pourraient posséder**,
- les informations supplémentaires nécessaires pour réidentifier les données sont conservées séparément, au sein de l'UE.

Anonymisation

Pas listée, mais RGPD ne s'applique plus au transfert

Chiffrement

- cryptage fort
- avant la transmission,
- clés gérées au sein de l'UE

  → utiliser une méthode de chiffrement qui permet encore la recherche parmi les données ?? (ex : hébergement cloud)

Notre fil conducteur

- La **société pharmaceutique** Harold Goldfarb est basée en **Australie**.
- Le projet de contrat d'étude prévoit que la société a accès aux données des patients sujets de l'étude **en clair.**

Les données synthétiques ?

Avis 3/2020 du CEPD sur la stratégie européenne pour les données :

- « Technologiquement prometteuses »
- « Pourraient faciliter l'accès aux données d'apprentissage pour l'apprentissage automatique »
- « des questions restent ouvertes (…) pour atténuer les risques en matière de protection des données »

Nous vous remercions pour votre attention

Avez-vous pensé à notre service d'abonnement mensuel à l'analyse de la jurisprudence de l'APD et de la Cour des marchés ?
Renseignements : f.coton@lexing.be