LES ENJEUX DE L'ANONYMAT DES DONNÉES MÉDICALES

Tessa Limbach

SOMMAIRE

Comment concilier l’exploitation des données de santé pour l’intérêt commun avec la garantie du respect des droits civiques et la protection des données ?

LES RISQUES

3 enjeux de la donnée de santé :

- intégrité
- confidentialité
- traçabilité

Les risques associés à la divulgation de ces données :

- Usurpation d'identité
- Non-respect des droits civiques
- Arnaques
- Ventes illégales
- Chantage
- Utilisation par des personnes mal intentionnées
- ...

LES DÉFIS

A) L'ANONYMAT

Une donnée anonyme est une donnée par laquelle il est impossible de remonter à l'individu physique auquel elle se rapporte.

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

Pseudo-anonymisation (temporaire) : patient en danger (violence, maltraitance, ...), personnalités célèbres, ...

Anonymat (irréversible) : accouchements sous X, prévention et dépistage des maladies sexuellement transmissibles, ...

- **Loi Informatique et Libertés**

Le traitement des données de santé est **interdit sauf exception particulière l'autorisant.**

- **Code de la santé**

- **RGPD**

L'article 9 § 2 du RGPD établit une liste d'exceptions permettant le traitement des données de santé :

- Consentement explicite
- Exécution de l'exercice des droits du responsable du traitement
- Sauvegarde des intérêts vitaux de la personne
- Défense d'un droit en justice
- Motifs "d'intérêt public importants"
- Médecine préventive ou médecine du travail
- Motifs d'intérêt public dans la santé publique
- Recherche scientifique ou historique ou statistique

SNIIRAM

Base nationale de données médico-administratives.

Plateformes de données de santé (GIP) :

SNDS

Le SNDS a été créé par la loi n° 2016-41 du 26 janvier 2016 de modernisation de notre système de santé.
Géré par la Caisse Nationale de l'Assurance Maladie (Cnam), il devait permettre de chaîner les données :

- de l'Assurance Maladie (SNIIRAM) ;
- des hôpitaux ;
- des causes médicales de décès
- relatives au handicap

SNDS
Système national des données de santé

CASD

Le CASD est un groupement d'intérêt public rassemblant l'État représenté par l'INSEE, le GENES, le CNRS, l'École polytechnique, HEC Paris et la Banque de France. Il a pour objet de mettre en œuvre des services d'accès sécurisé pour les données confidentielles à des fins non lucratives de recherche, d'étude, d'évaluation ou d'innovation.

CASD
CENTRE D'ACCÈS SÉCURISÉ AUX DONNÉES

LES DÉFIS

B) RISQUES DE RÉ-IDENTIFICATION

Le risque dépend :

- de la probabilité de réalisation de l'événement
- de l'impact

Risque = probabilité X impact

Cas de failles :

PREMERA | BLUE CROSS

Méthodes de dé-anonymisation

- Corrélation de données : croiser des données anonymisées avec des sources de données disponibles publiquement
- Utiliser une combinaison d'attributs
- Attaques en testant des combinaisons
- Rétroingénierie et détection de motifs
- Analyse temporelle
- ...

LES DÉFIS

C) RÉGLEMENTATIONS COMPLEXES

- **France**

Loi Informatique et Libertés

RGPD

Loi de Modernisation du Système de Santé

- **Union Européenne**

Directive sur la Protection des Données de Santé

RGPD

Autorités de Protection des Données (APD)

- **International**

Principes directeurs de l'OCDE sur la protection de la vie privée et les flux transfontières de données personnelles

Convention d'Oviedo

Divergences : Consentement - Anonymisation - Gestion des autorisations et des sanctions - ...

LA PLATEFORME HDH

A) CRÉATION ET FONCTIONNEMENT

2019 : l'article 41 de la loi relative à la transformation du système de santé est promulguée définit la Plateforme de données de santé (Health Data Hub).

Objectifs : croiser les bases de données de santé et faciliter leurs usages par les équipes de recherche et de développement.

Centralisation des masses de données de santé publique pour servir des projets d'intelligence artificielle.

Contexte :

- Plan "Intelligence Artificielle" lancé en 2019 par le Président de la République
- Volonté de développer le potentiel de l'IA dans la santé
- Rapport Villani : la France doit prendre des mesures lui permettant de mieux exploiter la masse d'informations médicales à disposition afin de favoriser les possibilités de l'IA, en facilitant un accès des données aux acteurs privés.

LA PLATEFORME HDH

B) PROBLÈMES ET RÉACTIONS

- **Confidentialité et protection des données**

Choix d'utiliser le cloud AZURE de **Microsoft** (filiale irlandaise de la société américaine) pour stocker et exploiter les données de santé de la population française.

Problème lié au **CLOUD Act** : loi qui permet à l'administration américaine de contraindre les fournisseurs de services américains à fournir les données stockées sur leurs serveurs

Forte critique des politiques, syndicats, associations ...

Arrêt SCHREMS II : invalide le Privacy Shield

- **Transparence**

- **Gouvernance**

Le HDH indique que les données sont placées sous sa responsabilité juridique, et qu'il « est responsable du traitement relatif à leur stockage, leur organisation et leur mise à disposition » : donc les responsables des sources de données n'auront plus la maîtrise des conditions de mise à disposition.

- **Consentement éclairé**

- **Indépendance du CESREES**

Organisation potentiellement source de conflits d'intérêt.

STRATÉGIES POUR GARANTIR L'ANONYMAT

- **Pseudonymisation**

Pseudonymat = une donnée nominative pour laquelle on a remplacé, dans le fichier source, l'identité de la personne par un code, la correspondance entre le code et l'identité réelle étant stockée à part dans un second fichier.

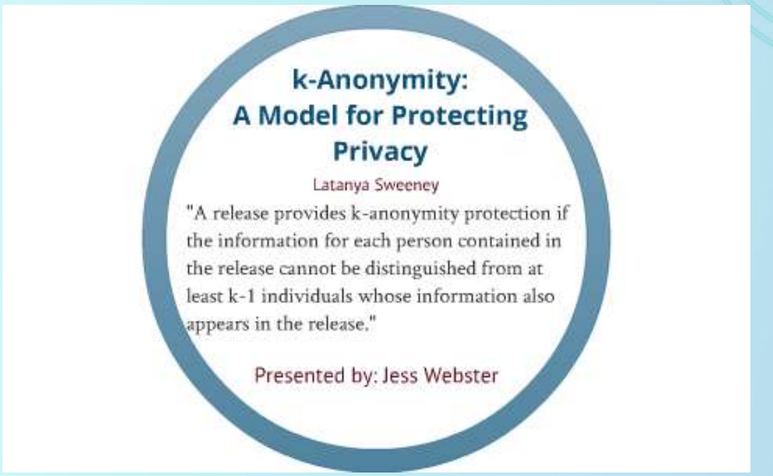
2 méthodes :

- Identifiant initial remplacé par une numérotation inrémentale
- Utiliser une fonction de hachage sur l'identifiant initial

- **K-anonymat et l-diversité**

K-anonymisé = pour toute clé d'identification, il existe au moins k individus indistingables du point de vue des variables quasi-identifiantes

l-diversité = pour chaque clé d'identification c, il existe au moins l modalités représentées pour chaque variable sensible.



STRATÉGIES POUR GARANTIR L’ANONYMAT

- **Cryptographie**

Chiffrement homomorphique : effectuer des calculs sur des données chiffrées sans les déchiffrer.

Chiffrement fonctionnel : permet de calculer une fonction des données ‘ex. moyenne), sans avoir accès aux données elles-mêmes

- **Agrégation**

Combinaison de plusieurs enregistrements en un seul ensemble de données pour supprimer les informations au niveau individuel.

- **Confidentialité différentielle**

Introduire du bruit ou du caractère aléatoire dans les données, ce qui rend difficile de déterminer si les données d'un individu sont incluses dans l'analyse.

Quand un utilisateur du fichier fait une requête, la réponse est bruitée. Le bruit ajouté est calculé pour que l’individu rentrant en compte dans le calcul ne puisse être identifié.

STRATÉGIES POUR GARANTIR L'ANONYMAT

- **Encadrement du NIR**

Le NIR ou numéro de sécurité sociale est attribué à chaque personne à sa naissance sur la base d'éléments d'état civil transmis par les mairies à l'INSEE.

Décret "cadre NIR" (2019) : établit les règles strictes concernant l'utilisation des données associées au NIR. Il énonce les mesures de sécurité à mettre en place pour prévenir tout accès non autorisé ou divulgation inappropriée.

- **Renforcement des sanctions**

- Sanctions Financières (amendes du RGPD pouvant atteindre 4% du CA mondial annuel d'une entreprise ou 20 millions d'euros) ;
- Sanctions si aucune alerte n'est donnée en cas de violation des données ;
- Réparation des dommages
- Sanctions rendues publiques
- ...

NÉCESSITÉ DE TROUVER
UN ÉQUILIBRE ENTRE PROGRÈS MÉDICAUX ET PROTECTION DE LA VIE PRIVÉE

Sensibilisation des acteurs médicaux