République Algérienne Démocratique et Populaire Université Abou Bakr Belkaid– Tlemcen
Faculté des Sciences Département d'Informatique

Mémoire de fin d'études

pour l'obtention du diplôme de Master en Informatique

Option: Réseaux et Systèmes distribués [R.S.D]

Thème

Mise en place d'une application de E-Santé assurant la protection de la vie privée des patients atteints de la COVID-19

Réalisé par :

Mr Bouziani Kheir Eddine

Et

Mr Brahimi Imad Eddine

Présenté le devant le jury composé de :

M. MANA Mohammed	Président
Mme AMRAOUI Asma	Examinatrice
Mme LABRAOUI Nabila	Encadreur
Mme SAIDI Hafida	Co-Encadreur

Année universitaire : 2021-2022

Remerciements

On remercie d'abord la grâce d'Allah, de nous avoir guidés et éclairés sur la bonne voie du savoir pour continuer ce travail et atteindre nos objectifs tracés puis on remercie nos parents qui nous ont soutenus et encouragés durant tout notre cursus universitaire que dieu les protège.

Avant tout, on tient à remercier notre encadreur Mme LABRAOUI Nabila et notre Co-encadreur Mme SAIDI Hafida pour leur soutient, leur aide, leurs conseils précieux et surtout leur patience avec nous durant toutes ces années de formation. Leurs disponibilités, leurs orientations précises et leurs qualités humaines et professionnelles ont contribué, majoritairement, à l'accomplissement de ce travail.

On tient, également, à remercier les membres du jury pour nous avoir honorés en acceptant de juger ce travail.

Je remercie M. MANA Mohammed d'avoir accepté de présider le jury, ainsi que Mme AMRAOUI Asma d'avoir accepté d'évaluer notre projet.

On remercie vivement tous nos enseignants du département d'informatique pour leurs efforts durant cette formation.

On n'oublie pas de remercier tous les étudiants de notre promo licence et master, Ils étaient très gentils et nous ont respectés et encouragés durant notre formation, on leur souhaite un bon avenir et une bonne continuité dans leurs vies.

On remercie toutes nos familles, nos amis et tous ceux qui ont contribué de près ou de loin à accomplir ce travail.

Table des matières

Liste de figures

Chapitre 1 :

Chapitre 2 :

Chapitre 3 :

Listes des tableaux

Chapitre 1 :

Chapitre 3 :

Introduction générale

L'e-santé (ou santé numérique) fait référence à « l'application des technologies de l'information et de la communication (TIC) à l'ensemble des activités en rapport avec la santé ». Le numérique est au cœur de l'innovation tant de la recherche clinique que dans la prise en charge et l'accompagnement des patients. Désormais des solutions évolutives et innovantes sont apparus tels que les « Smart Hopitaux » dans lequels les dossiers médicaux des patients sont numérisés et partagés par plusieurs entités telles que les médecins, les chercheurs, les analystes, etc.

Ces dernières années, le COVID-19 était le sujet principal de tous les médias et réseaux sociaux. Les personnes atteintes de cette pandémie partagent leurs informations confidentielles avec leurs médecins, ces informations peuvent être leurs maladies chroniques, leurs états familiales, leurs handicapes … etc.

D'un autre côté, des personnes, des entreprises et organisations malveillantes recherchent des informations utiles et ne vont pas laisser passer une telle chance, qui consiste à cueillir les informations de ces patients et les utiliser pour leurs intérêts, même si les données collectées peuvent être sous forme de statistiques, ils peuvent comparer entre les bases de données, et utiliser toute forme de méthodes pour en déduire « qui est atteint de telle maladie ? ».

C'est là qu'intervienne la confidentialité différentielle ! elle sert à ajouter à ces statistiques un certain « bruit », de sorte que les utilisateurs ne peuvent pas identifier des points de données individuels, par conséquent, La confidentialité personnelle est protégée avec un impact limité sur l'exactitude des données.

Le but de notre projet de fin d'études est d'appliquer la confidentialité différentielle sur une application que nous avons développé nous-mêmes et qui est un système d'information qui regroupe les médecins les patients, les analystes et chercheurs, et qui consiste à protéger la vie privée des patients tout en ajoutant un « bruit » aux données afin que d'autres utilisateurs comme l'analyste ou le chercheur ne puissent profiter de ces données.

Nous avons introduit dans l'application le concept du contrôle d'accès basé sur les rôles, qui est un moyen de limiter les faits et gestes de chaque utilisateur en fonction de son rôle, et cela pour garantir la sécurité de notre application, et pour renforcer sa crédibilité, étant donné que le but ultime de l'application est la sécurité des utilisateurs.

Dans ce mémoire nous allons détailler encore plus sur la confidentialité différentielle et les techniques d'anonymat, ainsi que sur le contrôle d'accès basé sur les rôles, sans oublier de donner les points à savoir sur notre application.

Structure du mémoire :

Notre mémoire est organisé en deux partie suivis d'une conclusion générale et perspectives.

1) La partie 1 : « état de l'art » qui comprend le chapitre 1 et le chapitre 2

- **Le chapitre 1** : Dans ce chapitre nous allons parler de E-Santé, et expliquer le concept de la vie privée et comment la préserver des attaques et des intrusions malveillantes a travers les différentes méthodes d'anonymisation.

- **Le chapitre 2** : Dans ce chapitre nous dirigerons notre recherche non seulement sur la confidentialité différentielle, mais aussi sur le contrôle d'accès basé sur les rôles (RBAC), et la combinaison des deux notions.

2) La partie 2 : « contribution » qui comprend le chapitre 3

- **Le chapitre 3 :** Dans ce chapitre nous allons donner les fondements de notre application, et cela en trois parties, la première partie qui est la conception, qui concerne les diagrammes UML (cas d'utilisation, séquence et classe), la deuxième partie qui est l'implémentation, c'est le mode de fonctionnement de notre application, et la troisième partie c'est l'évaluation des performances de notre application.

Résumé

La confidentialité différentielle est une propriété d'anonymisation qui consiste à ajouter un « bruit » ou une composante aléatoire au données personnelles, ce qui empêche les utilisateurs d'identifier les points de données individuelles, elle vise à éviter qu'un utilisateur puisse produire un nombre indéfini de rapports pour révéler des données sensibles.

Le but de notre projet de fin d'études est d'intégrer la confidentialité différentielle à une application E-Santé visant à protéger la vie privée des patients atteints de COVID-19, tout en appliquant le concept du contrôle d'accès basé sur les rôles (RBAC), où chaque utilisateur reçoit un rôle associé à des privilèges.

Mots clés : confidentialité différentielle, anonymisation, bruit, E-Santé, vie privée, contrôle d'accès basé sur les rôles (RBAC).

Abstract :

Differencial privacy is an anonymization property that consists of adding a « noise » or random component to personal data, which prevents users from identifying individual data points, it aims to prevent a user from producing an indefinite number of reports to reveal sensitive data.

The goal of our graduation project is to integrate differencial privacy into an E-Health application to protect the privacy of COVID-19 patients, while applying the Role-Based Acces Control (RBAC) concept, where each user receives a role associated with privileges.

Keywords : differencial privacy, anonymization, noise, E-Health, privacy, Role-Based Acces Control (RBAC).

ملخص:

الخصوصية التفاضلية هي خاصية لجعل الهوية مجهولة تتميز بإضافة "تشويش" أو عنصر عشوائي إلى البيانات الشخصية ، مما يمنع المستخدمين من تحديد البيانات الخاصة بالأفراد ، وتهدف إلى منع المستخدم من تجربة عدد غير محدود من المقارنات للكشف عن .البيانات الحساسة

الهدف من مشروعنا للتخرج هو دمج الخصوصية التفاضلية في تطبيق الصحة الإلكترونية لحماية خصوصية مرضى كوفيد-19 مع .تطبيق مفهوم التحكم في الولوج القائم على الأدوار ، حيث يتلقى كل مستخدم دورًا مرتبطًا بامتيازات ،

الكلمات المفتاحية : الخصوصية التفاضلية ، جعل الهوية مجهولة ، تشويش ، الصحة الإلكترونية ، خصوصية ، التحكم في الولوج .القائم على الأدوار

Chapitre 1 :

Respect de la Vie Privée

Sommaire :

Chapitre 1 : Respect de la vie privée

De nos jours, le digital s'insère progressivement dans les pratiques de l'ensemble des industries, notamment celles des industries de santé.

I.1. Introduction et Problématique

La E-santé englobe l'ensemble des innovations s'appuyant sur les technologies de l'information et de la communication pour la santé afin de collecter et générer des données. Elle impacte le monde médical à plusieurs niveaux allant de la prévention jusqu'à la guérison des patients.

D'un autre côté, le patient a une vie privée et ne veut pas forcément partager l'état de sa santé avec tout le monde, car il se peut que des gens ou des organisations malintentionnées, profitent des failles de ces informations pour leurs avantages et pour satisfaire leurs besoins, et il se peut même qu'ils nuisent à la vie de ces patients.

La vie privée des patients n'est plus protégée surtout depuis l'apparition du COVID-19, étant donné que les médecins, les chercheurs et les analystes ont accès à ces données, voire même des hackers qui profitent de cette période où la collecte d'informations est devenue monnaie courante, de ce fait les malades n'ont plus confiance aux applications, aux sites WEB où on leur demande leurs noms, leurs sexes et leurs âges ainsi que leurs maladies, et cela est devenu un problème à résoudre.

Et pour y remédier il faut se poser les questions suivantes :

- Comment protéger la vie privée des patients atteints de COVID-19 ?
- Existe-t-il une méthode fiable qui empêche les informations de ces patients de s'ébruiter tout en respectant leurs vies privées ? Et en quoi consiste cette méthode ?
- Comment peut-on mettre en œuvre cette méthode à travers une application informatique ?

Dans ce chapitre nous allons parler de E-Santé, de la pandémie du COVID-19 et expliquer le concept de la vie privée et comment la préserver des attaques et des intrusions malveillantes.

I.2. La E-santé

Alors que la population continue de vieillir et que le nombre de patients atteints de maladies chroniques augmente, la santé en ligne joue un rôle de plus en plus important.

Elle s'insère dans l'accompagnement des patients afin de permettre la réalisation de plusieurs objectifs : l'amélioration du suivi, la personnalisation et l'innovation des soins.

L'objectif est de réduire les inefficacités du système de santé et de réduire les coûts [10].

I.2.1. Définition de la E-Santé

La E-santé (ou santé connectée) consiste en l'utilisation des nouvelles technologies pour améliorer la santé des patients. Ces dispositifs facilitent l'accès aux soins et permettent à leurs utilisateurs de personnaliser les soins en termes préventifs ou médicaux.

La E-santé est un terme qui englobe à la fois la télémédecine (la télémédecine regroupe les pratiques médicales permises ou facilitées par les télécommunications), la télésanté (consultation à distance) et la m-santé (santé mobile).

La téléassistance vient en complément des solutions de santé E-santé, en permettant aux personnes de vivre plus sereinement, que ce soit à domicile ou en établissement [11].

I.2.2. L'importance de la E-santé

Dans les systèmes de E-santé, les services de santé combinent innovations technologiques et systèmes d'information pour améliorer la qualité des soins grâce à une meilleure gestion de l'information et une meilleure collaboration entre le personnel soignant, les hôpitaux et les centres médicaux. Quant à la téléassistance, elle permet de mieux protéger les personnes fragilisées et les seniors vivant seuls ou en EHPAD.

Le nouveau système est conçu pour améliorer la collaboration, l'efficacité et la qualité des soins dans de nombreux domaines de la santé. Le système permet également un échange d'informations simple, rapide et transparent. La cybersanté est une approche à multiples facettes de la prestation et de la gestion des soins de santé sociaux qui utilise les technologies de l'information pour fournir des soins de santé plus efficacement. Il permet la collaboration et l'échange d'informations à un niveau interdisciplinaire [11].

I.2.3. Les domaines d'application

Ces innovations ouvrent de nouveaux champs d'application médicale :

- Prévention primaire (c'est le domaine des applications et des assistants personnels)
- Dossier patient informatisé (dossier informatique rassemblant les données médicales de patients)
- Logiciel de gestion de cabinet médical
- Santé mobile (en relation avec les smartphones)
- Téléconsultation (consultation à distance)
- Téléassistance
- Systèmes d'information hospitaliers
- Aide à la décision (aide à la stratégie thérapeutique)
- Chirurgie assistée par ordinateur [11].

I.2.4. L’Internet des Objets (IOT) dans le milieu médical

Le développement d’un système de E-santé rend la pratique médicale plus efficace et permet aux prestataires médicaux et aux patients de mieux contrôler leurs états de santé.

L’usage d’objets IoT (Internet of things ou internet des objets) permet aussi de surveiller la santé des patients à domicile, ce qui augmente la commodité et l’efficacité pour les patients, les médecins et les hôpitaux.

Par exemple, les objets connectés tels que les lecteurs de glycémie permettent de mieux contrôler les objectifs glycémiques d'une personne de manière personnalisée, et d'avoir un journal de suivi personnalisable sur un smartphone avec les doses d'insuline injectées, les repas (par période horaire), l'activité.

Comme pour le diagnostic, la E-santé peut être appliquée au niveau préventif en mettant systématiquement en relation un patient avec son médecin de premier recours pour obtenir des informations opportunes sur son état et pour faire des recommandations de traitement rapides si nécessaire.

Cela n’aurait pas été possible avec les pratiques traditionnelles où chaque visite aurait pu nécessiter une demande d’information séparée de la part de différents services [11].

I.3. COVID-19 et technologie

En janvier 2020 le monde a été dévasté par le nouveau virus qui est apparu en premier lieu en chine, et a traversé le monde entier en infectant une grande partie de la population mondiale.

I.3.1. Définition de la pandémie

La maladie du coronavirus (COVID19) est une maladie infectieuse due au virus SARS-CoV-2.

La plupart des personnes infectées par le virus ont une maladie respiratoire légère à modérée et se rétablissent sans traitement spécial. Cependant, certaines personnes sont très malades et nécessitent des soins médicaux. Les personnes âgées et les personnes souffrant de problèmes médicaux sous-jacents tels que les maladies cardiovasculaires, le diabète, les maladies respiratoires chroniques ou le cancer sont plus à risque de maladie grave. N'importe qui, quel que soit son âge, peut tomber gravement malade ou mourir du COVID-19 [12].

I.3.2. Collecte et analyse de données sur la pandémie

La collecte des informations sur les cas atteints de COVID-19 peut s'avérer utile pour les hôpitaux, afin qu'ils puissent augmenter la qualité de soins, et faire un suivi sur l'évolution de la pandémie.

L'utilisation des mégadonnées pour prévoir les épidémies existait bien avant l'épidémie de COVID-19. L'une des premières grandes tentatives de Google a été d'utiliser son outil "Google Flue Trend" (GFT), qui est aujourd'hui retiré du service. Le principe de base est simple : avec une barre de recherche aussi largement utilisée que celle de Google, on pourrait s'attendre à ce qu'en cas de pandémie, de nombreuses personnes saisissent des mots clés spécifiques dans leur navigateur [1].

En comptant les cas atteints de la maladie, les centres hospitaliers envoient les chiffres au ministère de la santé, et ce dernier les envoie à son tour à des organisations mondiales comme l'OMS (Organisation mondiale de la santé), cette dernière utilise des moyens et outils techniques pour avoir des statistiques sur la pandémie à l'échelle mondiale.

L'OMS a annoncé une nouvelle approche pour améliorer l'accès aux données vitales : SCORE (survey), count (count), review (enable)) pour les données de santé. La pandémie de COVID-19 a mis en évidence le besoin urgent de données fiables et actualisées pour les actions sanitaires stratégiques.

I.4. Le Concept de la vie privée

Que voulons nous dire par « la vie privée » ?

La vie privée concerne toute information ou acte destiné à être individuel à une personne et à elle seule, et qui est privée au grand public, telle que l'identité de la personne, dossiers médicaux, conversations privées (mails, sms, etc.), photos et vidéos personnelles.

Afin de pouvoir préserver cette intimité et garder ces informations privées dans le monde du numérique, il existe plusieurs techniques tel que la cryptographie, qui se base sur le chiffrement des données à l'aide des clés, et pouvoir les échanger sur internet tout en les gardant privées [4].

Lorsque nous parlons de vie privée en ligne, il est facile de réduire le sujet à quelque chose qui peut être décrit comme « anti-écoute téléphonique », mais la vie privée est un concept beaucoup plus complexe que cela.

En général, on peut parler de sept types différents de vie privée, chaque type est important pour nos libertés civiles sous différents angles [17].

I.4.1. La vie privée du corps

La vie privée du corps signifie que votre corps est le vôtre, et les agents gouvernementaux ne peuvent pas l'examiner ou l'envahir sans votre consentement.

Le mot « envahir » ici n'implique pas de violence physique, c'est beaucoup plus banal que cela, par exemple prendre de force un échantillon de sang en perforant votre peau, est une invasion claire de votre corps. Examiner votre circulation sanguine pour des substances indésirables est également un exemple de violation de votre droit à la vie privée du corps. Par extension, vos pensées et vos émotions font également partie de l'intimité du corps avant qu'elles ne soient exprimées à quelqu'un, alors que vous pensez et ressentez simplement [17].

I.4.2. La vie privée de la correspondance

La vie privée de la correspondance est ce dont nous parlons habituellement lorsque nous discutons de la vie privée en ligne.

La lettre scellée analogique est un droit séculaire que les entreprises et les politiciens s'érodent alors que nous faisons le saut vers les communications numériques, et nous devons la défendre. La confidentialité de la correspondance signifie deux choses : la première, que vous avez le droit absolu de communiquer en privé avec qui vous choisissez, sans que personne ne vous surveille sans votre consentement, et la deuxième, le choix de la personne avec qui vous souhaiteriez communiquer (que ce soit une personne ou une machine) est également privé [17].

I.4.3. La vie privée des données

La vie privée des données est liée au saut vers le numérique. C'est l'intimité de votre journal intime dans votre maison qui n'est communiqué à personne. Vos photos, vos documents, vos données.

Aujourd'hui, les ordinateurs ont très peu de protection en matière de loi contre les perquisitions et les saisies, bien qu'ils soient beaucoup plus privés que le journal papier d'un adolescent, il faut y remédier.

En attendant, nous nous protégeons avec des utilitaires de chiffrement de disque complet tels que TrueCrypt et diverses saveurs de GNU/Linux qui ont un tel chiffrement de disque complet hors de la boîte [17].

I.4.4. La vie privée de la finance

Une renaissance intéressante est en cours de réalisation de la vie privée de la finance, alors que le bitcoin continue de s'implanter.

Autrefois, le secret bancaire signifiait que personne (même pas le gouvernement) n'avait été autorisé à avoir un aperçu de vos finances personnelles : richesse, dette, dépenses, revenus. Progressivement, le gouvernement a changé ces lois pour se donner un accès complet non seulement pour examiner votre économie, mais aussi pour la changer de force à volonté : saisir les impôts dus, par exemple, sans aucune action de votre part. La technologie Bitcoin a la promesse de restaurer le secret bancaire, mais sans aucune aide (ou consentement) des banques ou du gouvernement [17].

I.4.5. La vie privée de l’identité

Souvent négligé, la vie privée de l’identité est notre droit de nous occuper de notre vie quotidienne de manière anonyme.

Bien que nous ayons vu des demandes politiques de cartes d’identité de plus en plus fortes, et que la prétention de la sécurité aérienne empoisonne progressivement cette vie privée, l’une des plus fortes menaces à cette vie privée est une autre : c’est la prolifération des caméras de vidéosurveillance qui créent un réseau de caméras, qui, lorsqu’elles sont prises ensemble, est essentiellement capable d’enregistrer chaque pas extérieur de notre porte d’entrée [17].

I.4.6. La vie privée de localisation

Notre vie privée de localisation est le droit d'être où nous voulons sans qu'aucune partie du gouvernement ne sache à ce sujet.

Cette confidentialité a été essentiellement éliminée après 2001, car nos téléphones portables ont été transformés en appareils suivi par le gouvernement par le biais de lois sur la conservation des données, et nous devons reprendre cette confidentialité [17].

I.4.7. La vie privée du territoire

Notre vie privée du territoire est notre droit de ne pas avoir notre maison envahie par la force gouvernementale. Cela s’étendait au-delà de notre maison (nous prenons une petite partie de notre territoire avec nous alors que nous marchons et nous nous déplaçons : poches, sacs à main, le contenu de notre voiture) font tous partie de l’intimité du territoire [17].

I.5. Approches de protection de la vie privée

La vie privée et l'anonymat sont deux concepts proches mais différents. Ils sont tous les deux de plus en plus nécessaires pour remédier aux différents pièges et différents traçages présents de nos jours sur internet, que ce soit de manière légale ou non, il est important de comprendre pourquoi ils font partie intégrante de nos libertés civiles et pourquoi ils ne sont pas seulement bénéfiques pour l'individu, mais absolument critiques pour une société libre.

La vie privée est la capacité de garder ses informations personnelles pour soi-même, quel que soit leur impact sur la société [2].

Par exemple, un patient dans un hôpital a le droit d'exiger que son dossier médical ne soit être consulté par aucune personne à part son médecin ni même pas par sa sécurité sociale même si cette dernière le suspecte de fraude.

La vie privée est donc un concept décrivant les activités que nous gardons entièrement à nous-mêmes, ou à un groupe limité de personnes.

En revanche, l'anonymat à l'inverse du concept de la vie privée, nous souhaitons partager avec tout le monde (au public) une information mais à condition que notre identité ne soit pas révélée [2].

L'anonymat est donc une sorte de préservation de vie privée car nous souhaitons garder une information rien qu'à nous-mêmes qui est notre identité, nous pouvons donc dire que l'anonymat est inclus dans la vie privée [2].

Pour donner un exemple typique, lorsqu'une personne souhaite dénoncer un acte de criminalité mais qu'elle ne veut surtout pas que son identité soit divulguée pour des raisons de sécurité, la personne exige donc de garder son anonymat et c'est un choix privé qui rentre dans ses droits de préserver sa vie privée

L'opposition entre données personnellement identifiables et données anonymisées n'est pas absolue. C'est pourquoi il existe plusieurs méthodes d'anonymisation plus ou moins efficaces. Aujourd'hui, le "k-anonymat", la "l-diversité" ou le "secret différentiel" sont souvent utilisés, et la justification de ces techniques sera donnée ci-dessous. Différentes techniques sont jugées sur la sécurité qu'elles procurent et les résultats analytiques possibles qu'elles laissent derrière elles [3].

I.5.1. Le K-anonymat

Posons un ensemble de donnés T(A1, .. . ,An) et les quasi-identificateurs associés QIT, L'ensemble T est réputé k-anonyme si pour chaque quasi-identificateur QI appartenant à QIT, chaque séquence de tuples apparaît au moins k fois dans T[QI].

Le concept de k-anonymat émerge d'une généralisation de la dépersonnalisation par quasi-identificateurs. Étant donné que la plupart des ensembles de données n'ont pas la grande taille/le petit nombre requis de combinaisons de quasi-identifiants, une approche plus robuste a été développée. Le K-anonymat fait partie de ce que nous appelons une approche syntaxique (par opposition à une approche sémantique). Ces méthodes reposent sur une hiérarchisation des données, du plus général au plus spécifique, couramment appelée taxonomie (étude de l'évolution) [4].

Par exemple, une hiérarchisation très simple peut être représentée par la Figure II-5 : nous voyons que le Québec (élément général) peut être vu comme un ensemble de villes (Montréal, Québec, Rimouski), qui sont constituées de quartiers ou arrondissements. À la différence de cet exemple, les hiérarchies dans les méthodes d'anonymisation doivent être totales, c'est à dire que toutes les valeurs individuelles doivent s'y retrouver [4].

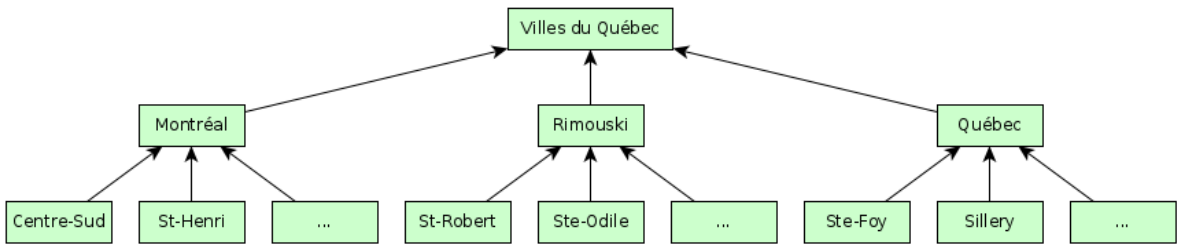


Figure I. 1 : Exemple de taxonomie : certaines villes et quartiers du Quebec [4]

Plus spécifiquement, cette méthodologie repose sur deux concepts permettant de fournir des garanties prouvables sur la confidentialité de l'ensemble de données résultant :

- La généralisation des paramètres, soit de remplacer une valeur spécifique par une valeur générale sans la falsifier ;
- La suppression pure et dure de ceux ne pouvant pas survivre à une généralisation satisfaisante [4].

I.5.2. La l-diversité

Une classe d'équivalence respecte la l-diversité s'il existe au moins l valeurs « bien représentées » pour l'attribut sensible. Une table respecte la l-diversité si chacune de ses classes d'équivalence respectent la l-diversité [5].

I.5.3. La t-proximité

Pour essayer de réduire encore l'information qui peut être observée directement, on introduit le modèle de la *t*-proximité, toujours à partir d'un regroupement de données en classes d'équivalences selon le processus du *k*-anonymat. Ce nouveau modèle s'appuie sur une connaissance globale de la distribution des données sensibles, c'est-à-dire pathologiques en l'occurrence, et tente de faire suivre au plus près cette distribution par les valeurs sensibles de la classe d'équivalence, évitant ainsi le raisonnement informatif provoqué par la question de l-diversité. Le facteur *t* que nous ne détaillons pas ici, indique dans quelle mesure on se démarque de la distribution globale [3].

Il y a plusieurs problèmes avec t-proximité, le plus important est probablement son côté pratique ! En fait, il semble évident d'exploiter des données k-anonymes voire l-diversifiées pour découvrir des corrélations entre des données appartenant à des quasi-identifiants et des données sensibles [3].

Cependant, le véritable objectif de la t-proximité est de réduire au maximum ces dépendances, puisque toutes les données sensibles se ressemblent pour chaque classe d'équivalence ! Ainsi, comme nous pouvons le voir dans le tableau I-3, la t-proximité répond surtout à la question : comment partitionner mes données afin que toutes les partitions aient une distribution similaire les unes par rapport aux autres ? [3].

I.5.4. La δ-présence

La δ-présence impose que la probabilité de présence d'un enregistrement soit dans un intervalle

δ = (δmin, δmax) prédéfini [5].

Le modèle de δ-présence a été mis en œuvre dans l'objectif de contrer les attaques par « lien de tables ». En effet, la publication de plusieurs tables anonymes par des éditeurs différents étant possible, on ne peut exclure la possibilité de rapprochement entre elles dès lors qu'elles partagent des valeurs de QI. Certains rapprochements peuvent mener à la divulgation de données sensibles [5].

I.5.5. Initiation à la confidentialité différentielle

Le problème de la confidentialité a deux objectifs contradictoires :

- Promesse de confidentialit**é :** Vos réponses resteront confidentielles et ne seront utilisées qu'à des fins statistiques.
- Utilité des données : Des rapports sont publiés à partir des données récoltées.

De temps à autre, des chercheurs ou le grand public ont accès à certaines données.

Alors comment peut-on s'assurer de respecter notre promesse de confidentialité tout en utilisant les données collectées ?

Une première solution intuitive consiste sur le fait d'anonymiser les données c'est-à-dire enlever toutes les variables qui pourraient permettre d'identifier directement le répondant (nom, adresse, …) et ce n'est malheureusement pas suffisant…

La deuxième solution est de ne publier que des données agrégées, encore une fois, ce n'est pas suffisant,

Pensez par exemple à un tableau avec une catégorie complètement vide, ou à une cellule avec un seul individu, ou encore à des données beaucoup plus compliquées…

Qu'est-ce qu'on fait alors ?

L'objectif c'est de maximiser l'utilité tout en minimisant le risque de divulgation [6].

Il y’a plusieurs approches que l’on puisse utiliser pour protéger les données :

- Limiter l'accès aux données.

  - Autoriser seulement les chercheurs d’une institution reconnue avec un projet de recherche sérieux à utiliser les données.

  - Établir une sorte de contrat qui interdit d’essayer d’identifier les individus .

- Confidentialiser les jeux de données, données agrégées (tableaux) ou sorties statistiques

  Avant la publication

  - À l’aide d’un logiciel statistique qui donne accès aux données [6].

Et il y’a deux types de méthodes pour confidentialiser les données :

- Méthodes de réduction (non-perturbatrices)
  - Masquer la valeur de certaines cellules dans un tableau de valeurs
  - Enlever certaines variables pour certains ou tous les individus
  - Partager seulement un échantillon des données
  - Combiner certaines catégories pour une variable catégorique
- Méthodes perturbatrices
  - Arrondir à la hausse ou à la baisse les valeurs extrêmes de certaines variables
  - Échanger les valeurs de certaines variables entre des répondants
  - Ajouter du bruit aléatoire aux données
  - Arrondir les fréquences dans un tableau
  - Créer des jeux de données complètement synthétiques [6]

Le problème de la confidentialité des données est important pour chacun

- En tant que citoyens partageant ses données,
- En tant que chercheur utilisant des jeux de données confidentiels.

Les solutions intuitives pour la protection de la confidentialité ne sont pas suffisantes. Définir/mesurer la protection de la confidentialité n'est pas simple.

La statistique est essentielle pour approcher ce problème [6].

I.5.6. Le pseudonymat

Le pseudonymat (ou pseudonymisation) consiste à supprimer les champs directement identifiants des enregistrements, et à rajouter à chaque enregistrement un nouveau champ, appelé pseudonyme dont la caractéristique est qu'il doit rendre impossible tout lien entre cette nouvelle valeur et la personne réelle.

Pour créer ce pseudonyme, on utilise souvent une fonction de hachage que l'on va appliquer à l'un des champs identifiants (par exemple le numéro de sécurité sociale), qui est un type de fonction particulier qui rend impossible (ou tout du moins extrêmement difficile) le fait de déduire la valeur initiale.

Dès lors, on voit que deux entités disposant des mêmes informations personnelles (identifiées par son numéro de sécurité sociale) peuvent partager ces données de façon anonyme en hachant cet identifiant. Il est également possible d'utiliser simplement une fonction aléatoire pour générer un identifiant unique pour chaque personne, mais cela ne résout pas tous les problèmes [3].

a) Avantage

Le gros avantage du pseudonymat est qu'il n'y a aucune limite sur le traitement subséquent des données. Tant que l'on traite des champs qui ne sont pas directement identifiés, on peut effectuer exactement les mêmes calculs que les bases de données non anonymes [3].

b) Exemple

Ainsi, on montre dans le tableau I-6 un exemple de calcul de la moyenne d'âge pour une pathologie donnée. L'utilisation de données pseudonymisées ne nuit pas à ce calcul [3].

Pseudonyme

Donnée sensible

ID	Age	CP	Sexe	Pathologie
1	75	75005	F	Cancer
2	40	75012	F	Grippe
3	12	78000	M	Grippe

Pathologie	MOY(age)
Cancer	75
Grippe	26

Tableau I. 1 : Pseudonymat et exemple de calcul [3]

c) Critique

Toutefois, le pseudonymat n'est pas reconnue comme un moyen d'anonymisation, car elle ne donne pas un niveau de protection suffisamment élevé : la combinaison d'autres champs peut permettre de retrouver l'individu concerné. Sweeney [3] l'a mis en évidence aux Etats-Unis en 2001 en croisant deux bases de données, une base de données médicale pseudonymisée et une liste électorale avec des données nominatives.

Le croisement a été effectué non pas sur des champs directement identifiants, mais sur un triplet de valeurs : code postal, date de naissance et sexe, qui est unique pour environ 80% de la population des Etats-Unis ! Elle a ainsi pu relier des données médicales à des individus (en l'occurrence le gouverneur de l'Etat) [3].

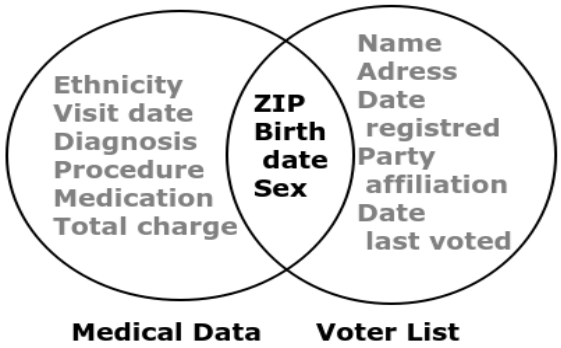


Figure I. 2 : Un exemple de recoupement d'une base anonyme [3]

I.6. Conclusion

La protection de la vie privée a toujours été menacée et l'est encore plus aujourd'hui, avec l'augmentation de l'influence des réseaux sociaux, la baisse de la contrainte morale et le développement des nouvelles techniques d'investigation [19].

Nous avons abordé dans ce chapitre les différentes techniques pour préserver la vie privée et nous avons donné un aperçu sur la confidentialité différentielle, qui nous sera très utile au prochain chapitre qui va être principalement dédié à la confidentialité différentielle, et au contrôle d'accès basé sur les rôles.

Chapitre 2 :

Intégration de la confidentialité différentielle au RBAC afin de protéger la vie privée

Sommaire :

Chapitre 2 : Intégration de la confidentialité différentielle au RBAC afin de protéger la vie privée

II.1 Introduction

La confidentialité différentielle ou intimité différentielle se démarque des tentatives précédentes de formalisation de la confidentialité, notamment du k-anonymat, par une perspective différente de l'enjeu, dans le chapitre précédent nous avons introduit les fondements de la confidentialité différentielle, mais dans ce chapitre nous approfondissons notre recherche non seulement sur la confidentialité différentielle mais aussi sur le contrôle d'accès basé sur les rôles.

II.2. La confidentialité différentielle

La définition la plus basique mathématique de la confidentialité différentielle est celle-ci : [6]

Une fonction randomisée K garantie la confidentialité différentielle de niveau ε si et seulement si pour tous jeux de données voisins D_1 et D_2 et pour tout S ∈ Image(K),

$$e^{-\varepsilon} \leq \frac{\Pr[K(D_1) \in S]}{\Pr[K(D_2) \in S]} \leq e^{\varepsilon} \qquad (2.1)$$

En effet il y'a d'autres définitions plus détaillées.

Nous dirons qu'une opération sur un ensemble de données BD(A) est différentiellement confidentielle si nous pouvons inférer environ la même quantité d'information sur un individu Y, qu'il soit présent ou non dans BD(A).

Plus formellement, nous pouvons dire :

ε-confidentialité différentielle : Un algorithme randomisé Ad est ε-différentiellement confidentiel si, pour tous les ensembles de données BD_1 et BD_2 différents d'au plus les données d'un seul individu et $\widehat{D} \subseteq$ Range(Ad) :

$\Pr[Ad(BD_1) \in \widehat{D}] \leq e^{\varepsilon} \bullet \Pr[Ad(BD_2) \in \widehat{D}]$ (2.2)

Les probabilités se trouvent sur celles de l'algorithme Ad.

Nous pouvons également relaxer la définition en ajoutant un paramètre additif δ qui nous permet d'ignorer les événements très rares.

(ε,δ)-confidentialité différentielle : Un algorithme randomisé Ad est (ε,δ)-différentiellement confidentiel si, pour tous les ensembles de données BD_1 et BD_2 différents d'au plus les données d'un seul individu et $\widehat{D} \subseteq$ Range(Ad) :

$$Pr[Ad(BD_1) \in \widehat{D}] \leq e^{\varepsilon} \bullet Pr[Ad(BD_2) \in \widehat{D}] + \delta \qquad (2.3)$$

Les probabilités se trouvent sur celles de l'algorithme Ad.

Les paramètres susmentionnés sont connus publiquement et leur valeur est inversement proportionnelle aux garanties de confidentialités suggérées par l'algorithme [4].

Cette définition peut paraître à prime abord comme une définition faible de la confidentialité.

En effet, notre but est idéalement de protéger les renseignements sensibles contenus dans notre ensemble de données.

En 1977, Dalenius a proposé la définition suivante, que nous appellerons confidentialité parfaite :

Aucune information sur un individu ne devrait être apprise en consultant la base de données BD(A) qui ne pourrait pas être apprise sans la consulter.

Dwork a prouvé que cette contrainte est impossible et cette inéquation a motivé la création de la définition de la confidentialité différentielle [4].

II.2.1. Principe

Une méthode très en vogue dans les milieux de la recherche en informatique depuis quelques années, car contrairement aux méthodes précédentes, elle est la seule à donner des garanties

formelles, c'est-à-dire des preuves mathématiques, sur la possibilité de borner les informations qu'on peut apprendre sur les individus [3].

La confidentialité différentielle est protégée en tenant attentivement compte des résultats des statistiques agrégées. Par exemple, un adversaire pourrait rétablir les données d'entrée en réalisant une analyse attentive des statistiques publiées. De même, si le public peut interroger une base de données sécurisée et qu'un tel accès lui permet de demander des statistiques simples sur un sous-ensemble de la base de données (moyenne, maximale, minimale, etc.), les attaquants pourraient abuser du système pour extraire les données d'entrée. La confidentialité différentielle réduit ce risque en ajoutant du "bruit" aux données d'entrée ou de sortie. À première vue, ce n'est qu'un exemple de la corruption des données utilisées dans les statistiques officielles depuis des décennies. La technique est affinée en employant une formulation mathématique rigoureuse de la confidentialité différentielle, qui permet d'évaluer précisément l'algorithme au point exact de l'échelle. « Protection de la vie privée – Utilité » au moyen d'un paramètre ε, ou epsilon [20].

Un algorithme porte le nom ε-différentiellement privé si l'exécution de l'algorithme dans deux bases de données dont seulement une entrée est différente produit des résultats qui diffèrent de moins de ε.

En termes simples, cela signifie qu'un adversaire utilisant les mêmes statistiques de différents sous-ensembles de la base de données ne peut déduire qu'une certaine quantité d'informations de la base de données liée par ε. En pratique, avant de publier des statistiques, déterminez le niveau de protection de la vie privée requis pour établir ε. On ajoute ensuite du « bruit aléatoire » aux données, jusqu'à ce que les algorithmes ou statistiques à calculer soient ε-différentiellement privés. Au moyen de la confidentialité différentielle, on garantit une meilleure protection des données de sortie tout en maximisant l'utilité [20].

Cette méthode introduit un échantillonnage des données vraies (avec une probabilité a), et une génération de données fictives avec une probabilité b >> a (mais ces données doivent naturellement rester réalistes...). Les garanties formelles sont cruciales pour quantifier le risque de ré-identification des tuples, il y a donc un enthousiasme pour cette approche. En fait, en examinant des ensembles de données anonymisés, les informations qui peuvent être obtenues pour savoir si un n-uplet est vrai ou faux sont doubles :

On n’est jamais sûr qu’un *n*-uplet soit vrai avec une probabilité supérieure à a, ni qu’il soit faux avec une probabilité inférieure à b [3].

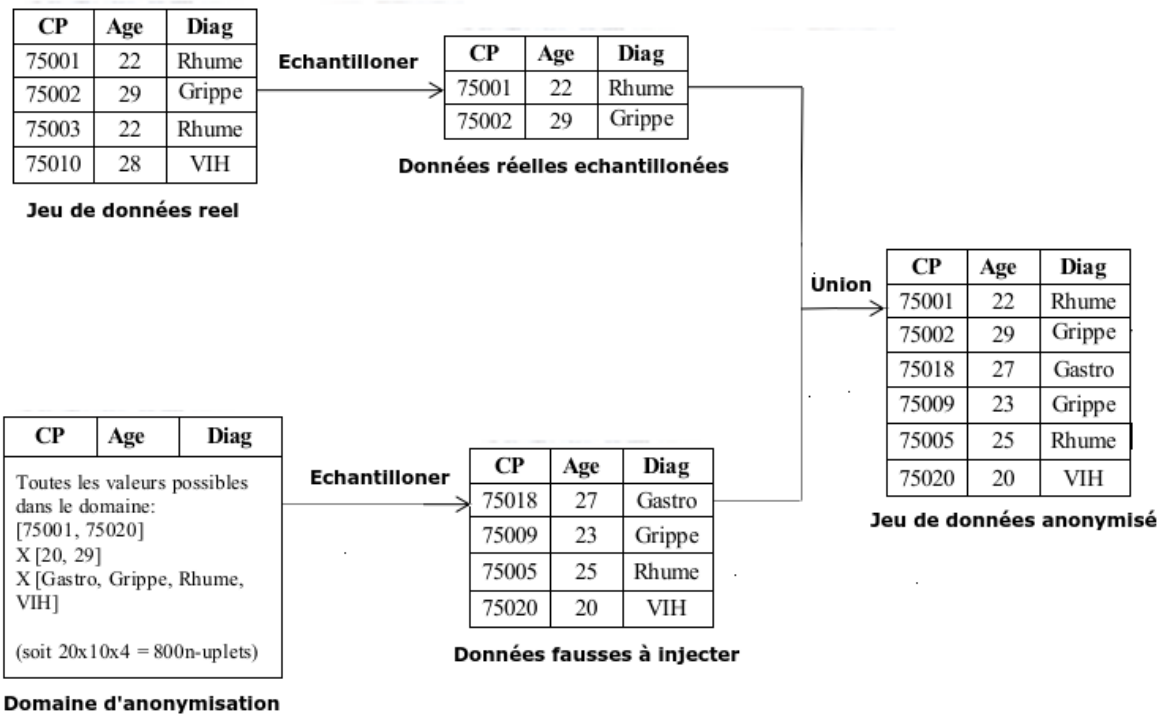


Figure II. 1 : Confidentialité différentielle [3]

La confidentialité différentielle oblige à calculer un estimateur d'un agrégat que l'on souhaite connaître [3].

Elle garantit aussi :

- De promettre aux répondants qu'une tierce personne ne pourra rien apprendre de plus sur eux qu'ils acceptent de participer à l'enquête que s'ils refusent.
- La protection rigoureusement mesurable de la confidentialité des données.
- De rendre le mécanisme de protection complètement public [6].

II.2.2. Exemple

Prenons l'exemple du calcul du nombre moyen de malades de la grippe par département, et supposons pour simplifier que les données fictives sont générées de manière équiprobable. On peut estimer le nombre total de malades de la grippe par la fonction suivante, dont l'objectif est de soustraire le bruit (connu) introduit :

$$Nb_{Rhume-estimé} = \frac{(Nb_{Rhume-anonyme} - \beta * Nb_{Rhume-domaine})}{\alpha} = \frac{2 - 200 * 0.005}{0.5} = 2 \qquad (2.4)$$

Le taux d'erreur peut également être estimé. Cependant, seules certaines fonctions d'agrégat peuvent calculer des erreurs bornées : moyenne, nombre total, etc.

En revanche, on voit bien que calculer la valeur maximale d'une donnée numérique ne fait pas sens [3].

En dehors de cette limitation, le principal problème pour obtenir une confidentialité différentielle est la rationalité des données virtuelles.

Ainsi, cette technique s'applique surtout lorsqu'on cherche à protéger des données de géolocalisation, où il est facile de générer des données fausses « plausibles », et où les fonctions qui peuvent être calculées à l'aide de cette technique d'anonymisation sont toujours utiles (en particulier la densité et la distance).

En revanche, comme on le voit sur l'exemple, il parait plus difficile d'exploiter cette méthode d'anonymat sur des données médicales [3].

II.2.3. Le budget de confidentialité

La confidentialité différentielle est un formalisme de protection robuste qui offre des garanties formelles quant à la protection des données.

Celui-ci impose que quelle que soit l'information qui peut être extraite à partir des données publiées, l'impact de la présence ou de l'absence d'un seul individu soit limité.

La confidentialité différentielle garantit notamment, que si un adversaire connaît toutes les informations de tous les enregistrements dans l'ensemble de données D mis à part un enregistrement, le résultat d'un algorithme probabiliste respectueux de la confidentialité différentielle ne doit pas fournir à l'adversaire trop d'informations additionnelles concernant l'enregistrement restant [7].

Un algorithme Probabiliste A satisfait la ε-confidentialité différentielle si pour deux ensembles de données voisins D_1 et D_2, et pour tout résultat possible O de A,

$$\Pr[A(D_1) = O] \leq e^{\varepsilon} \times \Pr[A(D_2) = O] . \qquad (2.5)$$

Le paramètre ε représente le budget de confidentialité.

Plus le epsilon est petit, plus le rapport $\frac{\Pr[A(D_1) = O]}{\Pr[A(D_2) = O]}$ est proche de 1. Ce qui revient à dire que les distributions de probabilité du résultat de l'algorithme A sur les ensembles voisins D_1 et D_2 sont approximativement égales.

Le ε représente une mesure relative, il est montré que pour la même valeur de ε, les garanties de protection de la vie privée offertes par la ε-confidentialité différentielle, varient en fonction du domaine d'attribut lui-même et de la requête prise en charge. En pratique, la question du choix de ε s'avère difficile et demeure aujourd'hui un défi [7].

II.2.4. Mécanismes d'ajout de bruit

On cite trois types de mécanismes d'ajout de bruit afin d'assurer la confidentialité différentielle :

a) Mécanisme Laplacien

Le mécanisme le plus répandu pour assurer la confidentialité différentielle est le mécanisme Laplacien, ce dernier fonctionne en ajoutant un bruit aléatoire à la réponse à une requête. En premier lieu, la vraie valeur de f (D) est calculée, où f est une fonction de requête et D c'est l'ensemble de données, ensuite un bruit aléatoire est ajouté à la fonction f (D) et la réponse

A (D) = f (D) + bruit est finalement retournée.

L'ampleur du bruit est choisie en fonction du plus grand changement que l'enregistrement peut causer à la sortie de la fonction de requête (Exemple : cela correspond à 1 pour une requête de dénombrement à travers D_1 et D_2), cette quantité définie par C.

Dwork l'a appelée sensibilité de la fonction [7].

Exemple : Sensibilité

Soient deux requêtes f_1 et f_2, telles que f_1 = 'Le nombre d'individus atteints d'hypertension' et f_2='La somme des âges des individus de l'ensemble de données'.

Supposons que age ∈ [0; 130], nous avons alors, $\Delta(f_1) = 1$ et $\Delta(f_2) = 130$ [7].

b) Mécanisme Exponentiel

Le mécanisme Laplacien a été mis au point pour les requêtes dont les résultats sont numériques.

Pour les requêtes dont les résultats ne sont pas numériques, ont proposé le mécanisme exponentiel. Celui-ci est basé sur une fonction d'utilité qui évalue l'utilité de chaque résultat possible à une requête, puis sélectionne une sortie t ∈ T qui est proche de l'optimum (au sens de la fonction d'utilité), tout en respectant la confidentialité différentielle [7].

Par exemple :

Soit la requête suivante, quelle-est la nationalité la plus courante dans l'ensemble de données D,

A partir de T = {Chinoise, Indienne, Américaine, Grecque}.

La fonction u(D,t) = nombre d'individus dans D qui possèdent la nationalité t, pourrait être envisagée comme fonction d'utilité.

La sensibilité serait donc Δu = 1, Le mécanisme retourne une nationalité qui est partagée par K individus avec une probabilité exp($\frac{\varepsilon K}{2}$). [7]

c) Mécanisme Gaussien

Dans le cas d'un bruit gaussien, nous considérons une version plus faible de la confidentialité différentielle qui dépend de deux paramètres ε et δ comme ceci :

$$\Pr[A(D_1) = O] \leq e^{\varepsilon} \times \Pr[A(D_2) = O] + \delta \qquad (2.6)$$

Dans certains cas particuliers, le responsable de traitement peut ne pas connaître la sensibilité Δ des données à bruiter. Ainsi, dans la situation où les données sont distribuées entre plusieurs entités, chacune d'entre elles n'a accès qu'à une fraction des données à traiter [8].

II.2.5. Confidentialité différentielle locale et globale

Le mécanisme de la confidentialité différentielle ajoute essentiellement du bruit (typiquement Gaussien ou Laplacien) aux données brutes pour atteindre un niveau quantifiable de confidentialité.

En connaissant ce niveau, on peut estimer la quantité d'informations maximale qui pourra être divulguée dans notre jeu de données [21].

Il existe deux méthodes principales :

a) Confidentialité différentielle locale

Le bruit est ajouté à chaque point de données individuel dans l'ensemble de données (soit par l'un des collaborateurs de l'entreprise après avoir obtenu les données, soit par l'individu lui-même avant de fournir les données à l'entreprise) [21].

Dans ce cas, chaque utilisateur applique l'algorithme différentiellement privé à ses propres données.

Ensuite, ils n'envoient leurs données à l'agrégateur qu'une fois qu'elles sont déjà anonymisées.

L'agrégateur n'a pas accès aux données réelles. Un agrégateur est un organisme ou un centre de recherche qui collecte et traite les données [22].

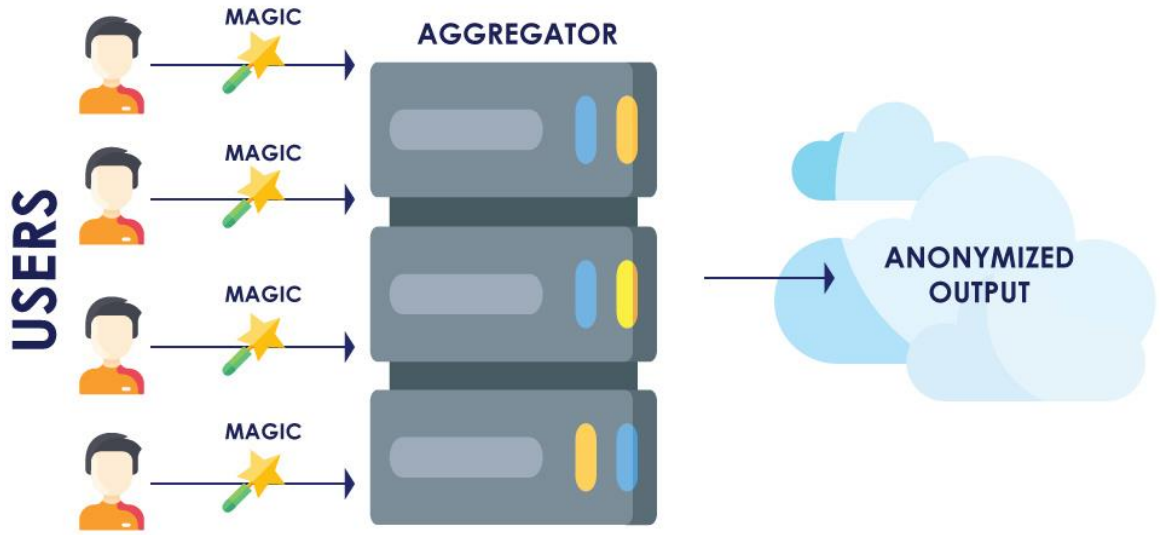


Figure II. 2 : Confidentialité différentielle locale [21]

b) Confidentialité différentielle globale

Le bruit requis pour protéger la vie privée des individus est ajouté à la sortie des requêtes effectuées sur les données brutes.

En général, la confidentialité différentielle globale peut produire des résultats plus précis que la confidentialité différentielle locale tout en maintenant le même niveau de confidentialité.

D'autre part, lors de l'utilisation de la confidentialité différentielle globale, la personne qui donne les données doit faire confiance à l'entité réceptrice pour ajouter le bruit nécessaire au maintien de sa confidentialité [21].

Dans ce modèle, il existe un agrégateur central. Chaque utilisateur envoie ses données sans bruit à cet agrégateur.

L'agrégateur prend ces données et les transforme avec un algorithme différentiellement privé [22].

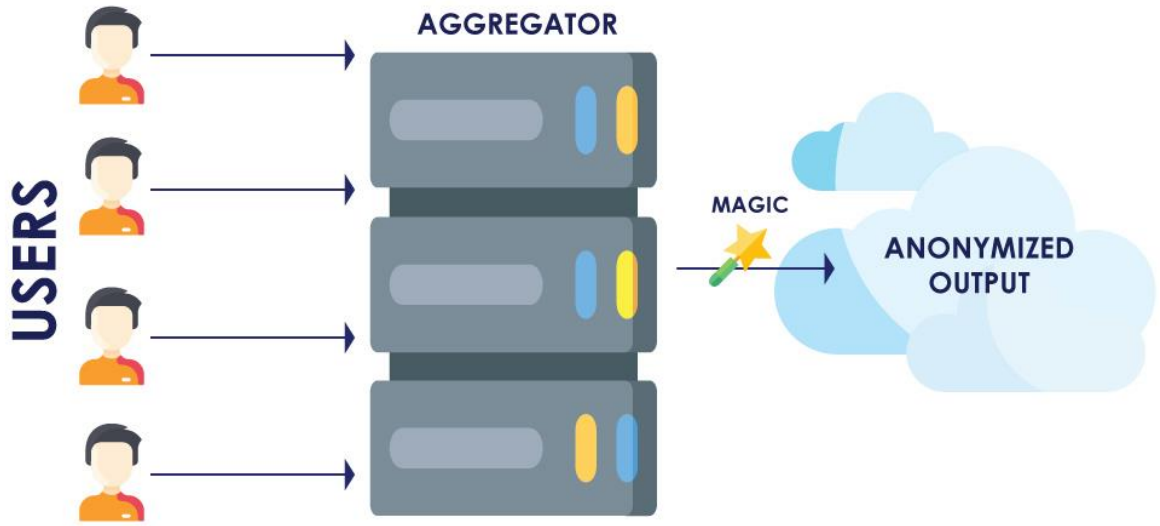


Figure II. 3 : Confidentialité différentielle globale [21]

II.3. Le contrôle d’accès base sur les rôles (RBAC)

Le contrôle d'accès basé sur les rôles (RBAC) est une méthode de sécurité d'accès basée sur le rôle d'un individu dans une entreprise.

Le contrôle d'accès basé sur les rôles est un moyen d'assurer la sécurité car il permet aux employés d'accéder uniquement aux informations dont ils ont besoin pour faire leur travail, tout en les empêchant d'accéder à d'autres informations qui ne les concernent pas. Les rôles des employés déterminent les autorisations qui leur sont accordées et garantissent que les employés de niveau inférieur ne peuvent pas accéder aux informations sensibles ou effectuer des tâches avancées [23].

II.3.1. Principe

Dans RBAC, il y a des règles :

1. Une personne doit se voir attribuer un certain rôle afin de mener une certaine action, appelée transaction.
2. Un utilisateur a besoin d'une autorisation de rôle pour être autorisé à détenir ce rôle.
3. L'autorisation de transaction permet à l'utilisateur d'effectuer certaines transactions.
4. La transaction doit être autorisée via l'appartenance au rôle. Les utilisateurs ne pourront pas effectuer de transactions autres que celles pour lesquelles ils sont autorisés [23].

Tous les accès sont contrôlés par des rôles attribués par l'utilisateur, qui sont un ensemble d'autorisations. Le rôle d'un employé détermine les autorisations qui lui sont accordées.

Par exemple, le PDG (Président Directeur Général) se verra attribuer le rôle de PDG et disposera des autorisations associées à ce rôle, tandis que l'administrateur réseau se verra attribuer le rôle d'administrateur réseau et disposera de toutes les autorisations associées à ce rôle. [23].

II.3.2. Fonctionnement

Avant d'appliquer le concept d'autorisation RBAC à une entreprise, les autorisations des rôles doivent être spécifiées avec autant de détails que possible.

Cela inclut le réglage fin des autorisations dans les domaines suivants :

- Droits de modification des données (lecture, lecture et écriture, accès complet)
- Droits d’accès aux applications de l’entreprise
- Autorisations dans les applications

Pour tirer pleinement parti du modèle RBAC, le développement des concepts de rôles et d'autorisations se démarque toujours. Les organisations déplacent toutes les fonctions des collaborateurs vers des rôles qui définissent les droits d'accès appropriés. Dans la deuxième étape, des rôles sont attribués aux collaborateurs en fonction des tâches [24].

Le contrôle d'accès basé sur les rôles permet à chaque utilisateur de se voir attribuer un ou plusieurs rôles. Les droits d'accès peuvent ainsi être attribués individuellement dans le modèle de rôle. Par conséquent, l'utilisateur peut obtenir l'autorisation d'effectuer sa tâche sans autres modifications.

Le RBAC est mis en œuvre et surveillé via un système de gestion des accès aux identités, ou IAM (système de gestion des identités) en abrégé.

Le système aide essentiellement les entreprises comptant un grand nombre d'employés lors de l'attribution, du contrôle et de la mise à jour de toutes les identités et des droits d'accès. L'octroi d'autorisations est appelé " Provisioning" et le retrait est appelé " De-Provisioning".

Une condition préalable à l'utilisation d'un tel système est l'établissement d'un concept de rôle unifié et standardisé [24].

Le contrôle d'accès basé sur les rôles repose sur une structure à trois niveaux d'utilisateurs, de rôles et de groupes.

Dans ce que l'on appelle le « role mining », les organisations définissent les rôles qui dépendent généralement de la structure organisationnelle de l'entreprise.

Enfin, chaque collaborateur se voit attribuer un ou plusieurs rôles, dont un ou plusieurs droits d'accès. Un ou plusieurs groupes sont associés à un rôle sans nécessairement être au même niveau.

Pour développer la notion de rôle, la méthode pyramidale s'impose la plupart du temps :

- **Le sommet :** autorisations pour tous les collaborateurs

  Le sommet définit les autorisations requises pour tous les employés de l'organisation, telles que l'accès à l'intranet, à la suite Office, au courrier électronique des clients, à l'ensemble du registre du réseau ou à l'enregistrement dans Active Directory [24].

- **Le deuxième niveau :** appartenance à un service

  Au sein d'une organisation, les employés d'un département sont responsables des activités dans le même domaine. Par conséquent, le service financier doit avoir accès au système ERP et aux disques durs du service, tandis que le service des ressources humaines doit avoir accès à toutes les données des employés. Les autorisations correspondantes sont attribuées à tous les employés d'un service.

- **Le troisième niveau :** les fonctions d'autres autorisations sont définies selon les fonctions des employés et des tâches qu'ils accomplissent [24].

II.3.3. Avantages et inconvénients

Sous certaines conditions, le contrôle d'accès basé sur les rôles est devenu un modèle de bonne pratique. Si les notions de rôles et d'habilitations sont définies et appliquées de manière obligatoire dans toute l'entreprise, le RBAC présente de nombreux avantages :

- **Flexibilité** : l'entreprise attribue seulement un ou plusieurs rôles à un collaborateur selon les besoins. Les modifications de la structure organisationnelle ou des autorisations sont rapidement communiquées à tous les employés à mesure que l'entreprise s'adapte aux rôles appropriés.
- **Faible charge administrative** : le RBAC rend obsolète l'attribution coûteuse d'autorisations individuelles.
- **Risque d'erreurs réduit** : les autorisations individuelles sont plus coûteuses et aussi plus sujettes aux erreurs que l'attribution d'autorisations d'accès à base de rôles.

- **Augmentation de l'efficacité** : en réduisant la charge et le risque d'erreurs, l'efficacité du système informatique et des autres collaborateurs augmente. Les modifications manuelles, le traitement des erreurs, les délais d'attente ainsi que les demandes individuelles de droits disparaissent.

- **Sécurité** : les droits d'accès sont définis exclusivement selon le concept de rôles, ce qui évite la sur-autorisation de collaborateurs individuels. Ceci correspond au principe du moindre privilège.

- **Transparence** : la désignation des rôles est le plus souvent aisément compréhensible, ce qui renforce la transparence et la compréhension par les utilisateurs [24].

Les points négatifs du contrôle d'accès basé sur les rôles sont :

- **Élaboration complexe** : le transfert des structures de l'organisation dans le modèle RBAC requiert beaucoup de travail.

- **Attribution temporaire** : si un utilisateur a besoin temporairement de droits d'accès étendus, il est plus facile d'oublier l'attribution avec le RBAC qu'avec une attribution individuelle de droits.

- **Utilisation** : pour les petites entreprises, l'élaboration et la maintenance des rôles nécessitent davantage de travail que la répartition des droits individuels. Par conséquent, le modèle RBAC n'est utilisé que pour un certain nombre de rôles et de collaborateurs. Et pourtant, même pour les grandes entreprises, le Role Based Access Control a pour inconvénient de multiplier les rôles. Si une entreprise a dix services et dix rôles, il en résulte déjà 100 groupes différents [24].

II.4. Préservation de la vie privée en se basant sur la confidentialité différentielle et le RBAC

Dans notre projet de fin d'étude, nous proposons une combinaison du contrôle d'accès basé sur les rôles et de confidentialité différentielle. La motivation derrière cette idée est de fournir plusieurs niveaux de précision pour les données d'entrée et les résultats par rapport aux différents utilisateurs ayant plusieurs rôles.

C'est-à-dire des réponses plus précises pour des rôles plus fiables, tandis que des réponses plus générales et bruyantes pour d'autres, en fonction du ou des rôles qu'un utilisateur a dans le système. Le système retourne normalement un seul résultat global qui ne peut en aucun cas révéler des informations sur un individu, sauf si le rôle qui a fait cette demande a un niveau élevé de sécurité

Dans le modèle du RBAC, les politiques ne seront pas modifiées dans le cas de réaffectation de personnes à différents rôles au sein de l'organisation, parce que les autorisations sont spécifiés en termes de rôles plutôt que de personnes.

Une autre motivation pour utiliser le RBAC est la hiérarchie des rôles, une façon de structurer les rôles à travers la relation d'héritage entre les rôles, par laquelle un rôle (souvent un rôle majeur de l'organisation) est autorisé à hériter des autorisations des rôles mineurs [9].

En fait, les hiérarchies des rôles peuvent être vues comme accumulation de permissions d'autres rôles. Par exemple, le rôle du médecin hérite des autorisations des rôles d'infirmière et de technicienne de laboratoire, car il s'agit de rôles mineurs. Cette composante du RBAC simplifie considérablement l'administration du contrôle d'accès en réutilisant la spécification d'un rôle par un rapport hiérarchique entre les rôles, ce qui évite le nécessité de spécifier de nouveau les autorisations pour chaque rôle [9].

II.4.1. Qui a accès à quel type de données ?

En fonction du ou des rôles qu'un utilisateur a au sein d'une organisation, il peut accéder à certains types de données et soumettre un nombre limité ou illimité de requêtes.

En fait, l'exactitude des réponses pour les requêtes des utilisateurs dépendent de leur rôle(s) dans le système. Donc, la quantité de bruit aléatoire ajouté au résultat pour obtenir une confidentialité différentielle et le nombre de questions répondues varie selon leurs rôles.

Une organisation doit d'abord définir différents rôles en termes de quelles informations peuvent être consultées par chaque rôle, puis attribuer des autorisations à chaque rôle.

Par la suite, les garanties de la confidentialité différentielle seront appliquées en fonction de ces rôles autorisations. La gestion des budgets de confidentialité est un problème administratif pour tous les systèmes de confidentialité différentielle.

Dans le cas de répartition inefficace du budget de confidentialité où se trouve un budget unique en matière de protection de la vie privée spécifié pour tous les rôles, un rôle peut consommer plus de budget qu'il en a réellement besoin ou mérité.

Par conséquent, pour obtenir une confidentialité différentielle avec RBAC, le propriétaire des données doit définir différents rôles en fonction du type de données qu'il possède [9], puis spécifier les paramètres suivants :

- Multiples valeurs de ε pour contrôler le niveau de confidentialité. Comme mentionné précédemment, un ε plus petit donne plus de confidentialité mais moins de précision. Par exemple, pour calculer la moyenne des bourses d'études sur un ensemble de données universitaires. Le fournisseur de données définit un plus grand ε pour des rôles plus fiables (par exemple, directeur du département) menant à ajouter moins de bruit à la sortie, et ainsi fournir un résultat plus précis.
- Multiples valeurs du budget de confidentialité pour assurer l'équité. Le même budget de confidentialité peut être attribué à tous les rôles tandis que le système soustrait différentes valeurs de ε selon différents rôles de ce budget après avoir soumis une requête. Une meilleure façon de préserver la vie privée pourrait envisager des valeurs différentes du budget en fonction des rôles [9].

Lorsqu'un demandeur envoie une demande pour accéder aux données, le système vérifie d'abord son rôle, et examiner ses niveaux d'accès pour décider comment répondre à sa demande.

Par exemple, un demandeur externe peut avoir un budget fixe et limité pour soumettre ses requêtes, tandis qu'un demandeur interne il a non seulement plus d'accès à tous les produits et données, mais aussi exécuter plus de requêtes (sous la forme de plus de budget de confidentialité) tout en recevant des réponses plus précises (sous forme de plus grande valeur d'epsilon).

Comme l'utilisateur peut jouer plusieurs rôles, le système applique l'une des situations suivantes après que l'utilisateur se connecte :

1- Il obtiendra toutes les missions associées à ses multiples rôles en même temps (RBAC de base).

2. Il obtiendra soit toutes les autorisations des rôles mineurs, soit héritera seulement du rôle mineur immédiat basé respectivement sur des hiérarchies générales ou limitées (RBAC hiérarchique).

3. Il n'obtiendra que des autorisations concernant le rôle par lequel il se connecte, et pendant cette session, les autorisations de ses autres rôles ne sont plus accessibles pour lui (RBAC restreint) [9].

II.5 Conclusion

Nous avons abordé dans ce chapitre les fondements de la confidentialité différentielle ains que le contrôle d'accès basé sur les rôles (RBAC). Nous verrons dans le chapitre suivant la conception de notre application.

Chapitre 3 :

Conception et implémentation

Sommaire :

Chapitre III : Conception et implémentation

Dans ce chapitre nous allons donner les fondements de notre application, et cela en trois parties, la première partie qui est la conception, où nous allons tracer les diagrammes UML (cas d'utilisation, séquence et classe), car c'est ce qui nous permet de structurer notre application, la deuxième partie qui est l'implémentation, c'est le mode de fonctionnement de notre application ainsi que les outils que nous avons utilisé dans ce projet, tandis que la troisiéme partie c'est l'évaluation des performances de notre application.

III.1. Conception

Le diagramme de cas d'utilisation va nous permettre d'attribuer à chaque acteur son rôle, le diagramme de séquence est la suite logique d'évènements qui se passent entre les acteurs dans l'application et le diagramme de classe c'est lui qui va nous aider à mettre en place la base de données de notre application

III.1.1. Diagrammes de cas d'utilisation

Il existe 4 acteurs concrets : l'utilisateur, l'admin, le patient, le médecin, le chercheur et l'analyste, et un acteur abstrait : l'utilisateur.

Chaque acteur concret est considéré comme utilisateur.

a) L'administrateur

L'admin comme tout autre utilisateur doit d'abord s'identifier pour accéder a ses fonctions, il peut créer un utilisateur sur commande (c'est-à-dire obtenir un nom d'utilisateur et un mot de passe seulement pour des personnes autorisées à accéder à l'application et qui lui ont demandé de leur donner un compte)

Il peut également supprimer un utilisateur, et aussi consulter les différentes listes d'utilisateurs.

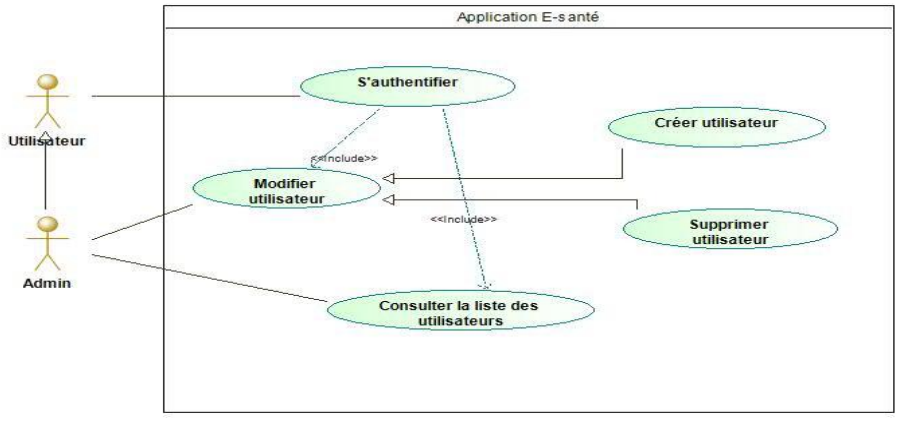


Figure III. 1 : Diagramme de cas d'utilisation d'administrateur

b) Le patient et le médecin

Le patient doit s'identifier avant de déposer son dossier médical, il peut aussi consulter ses informations personnelles et même les modifier, Le médecin, quant à lui, quand il se connecte, il peut consulter la liste de ses patients et consulter les dossiers médicaux des patients, il peut aussi modifier l'état (ou le statut) du patient, par exemple : de l'état « positif » à l'état « vacciné ».

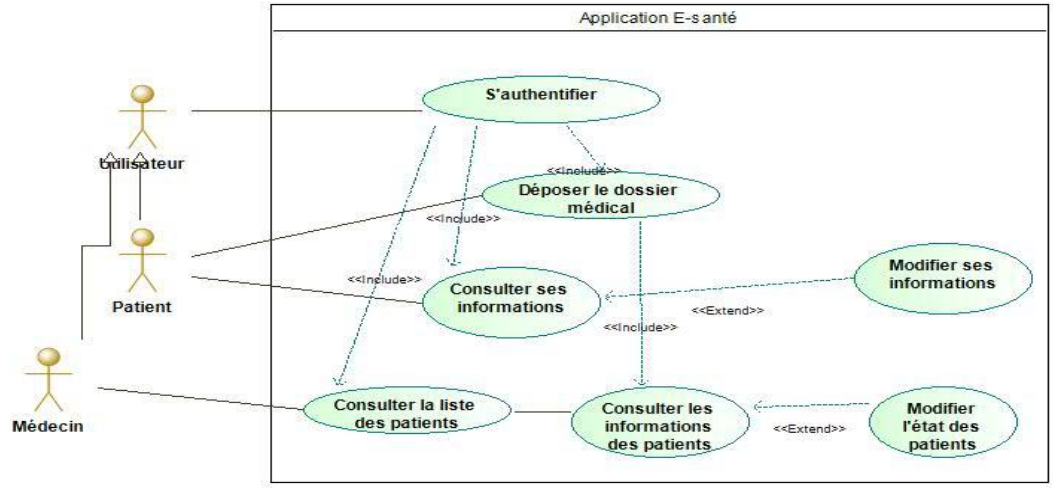


Figure III.2 : Diagramme de cas d'utilisation du patient et du médecin

c) Le chercheur et l'analyste

L'analyste et le chercheur, après identification sur l'application, ils peuvent tous les deux consulter les statistiques du COVID-19 affichées par l'application (qui sont basées sur les patients inscrits), la seule différence entre eux c'est que le chercheur peut accéder aux dossiers médicaux des patients mais sans lire les noms des patients concernés, et l'analyste peut consulter le nombre de patients vaccinés ou positifs mais sans pouvoir en déduire de qui il s'agit.

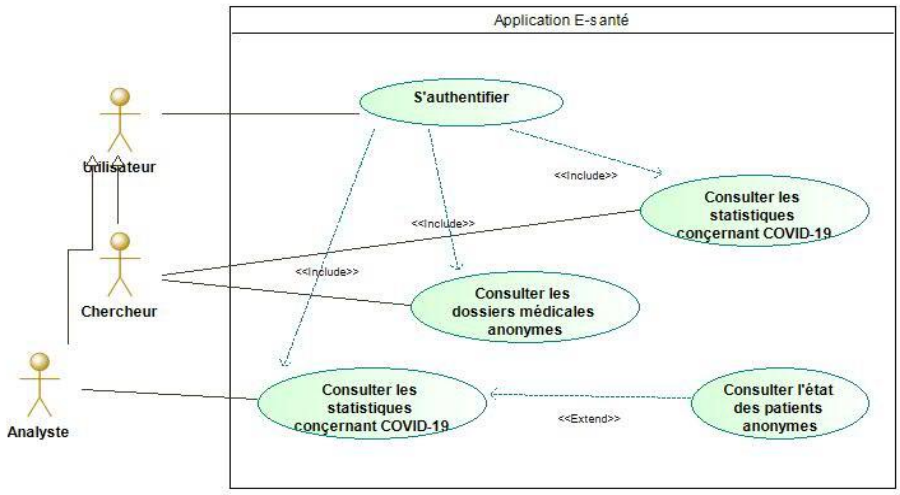


Figure III. 3 : Diagramme de cas d'utilisation du chercheur et de l'analyste

III.1.2. Diagrammes de séquence

On compte trois diagrammes de séquence, chaque diagramme représente un déroulement particulier, le diagramme de l'identification, le diagramme d'une personne qui demande un compte d'accès à l'admin et le diagramme du déroulement normal au sein de l'application.

a) S'identifier

Ce diagramme indique les scénarios possibles de se dérouler lors de l'identification.

Deux possibilités s'offrent à l'utilisateur qui souhaite se connecter :

1- il entre son login et son mot de passe dans l'interface d'authentification, son login et son mot de passe sont corrects d'après la base de données donc il permet d'accéder à l'application

2- le login et le mot de passe sont incorrects, donc l'accès est refusé.

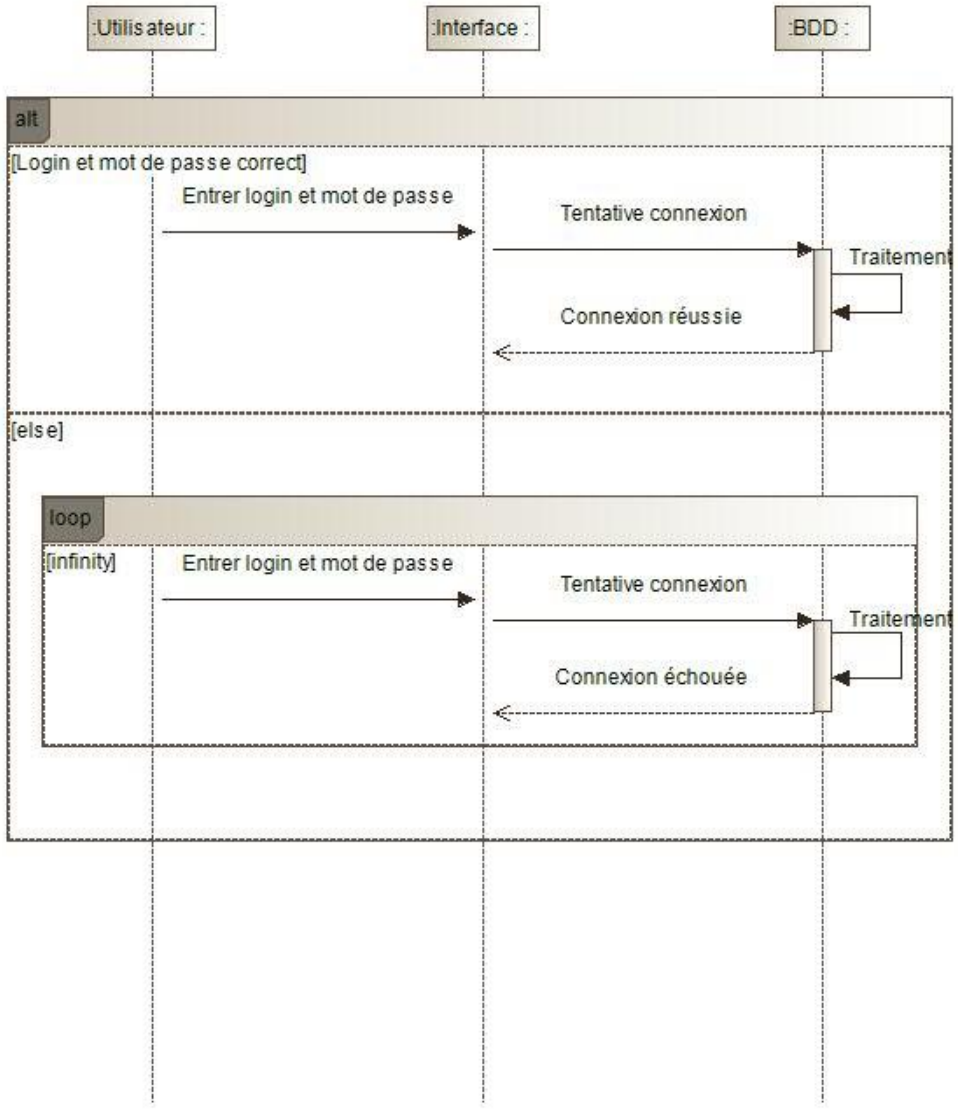


Figure III. 4 : Diagramme de séquence de l'identification

b) Demander un compte

Une personne souhaitant accéder à l'application doit demander à l'administrateur de lui donner un login et un mot de passe, l'admin vérifie si la personne en question est autorisée à accéder à l'application, si la personne est autorisée alors il lui crée un compte et lui donne son nom d'utilisateur et mot de passe pour qu'elle puisse se connecter, dans le cas contraire (accès non autorisée), la personne verra sa demande refusée.

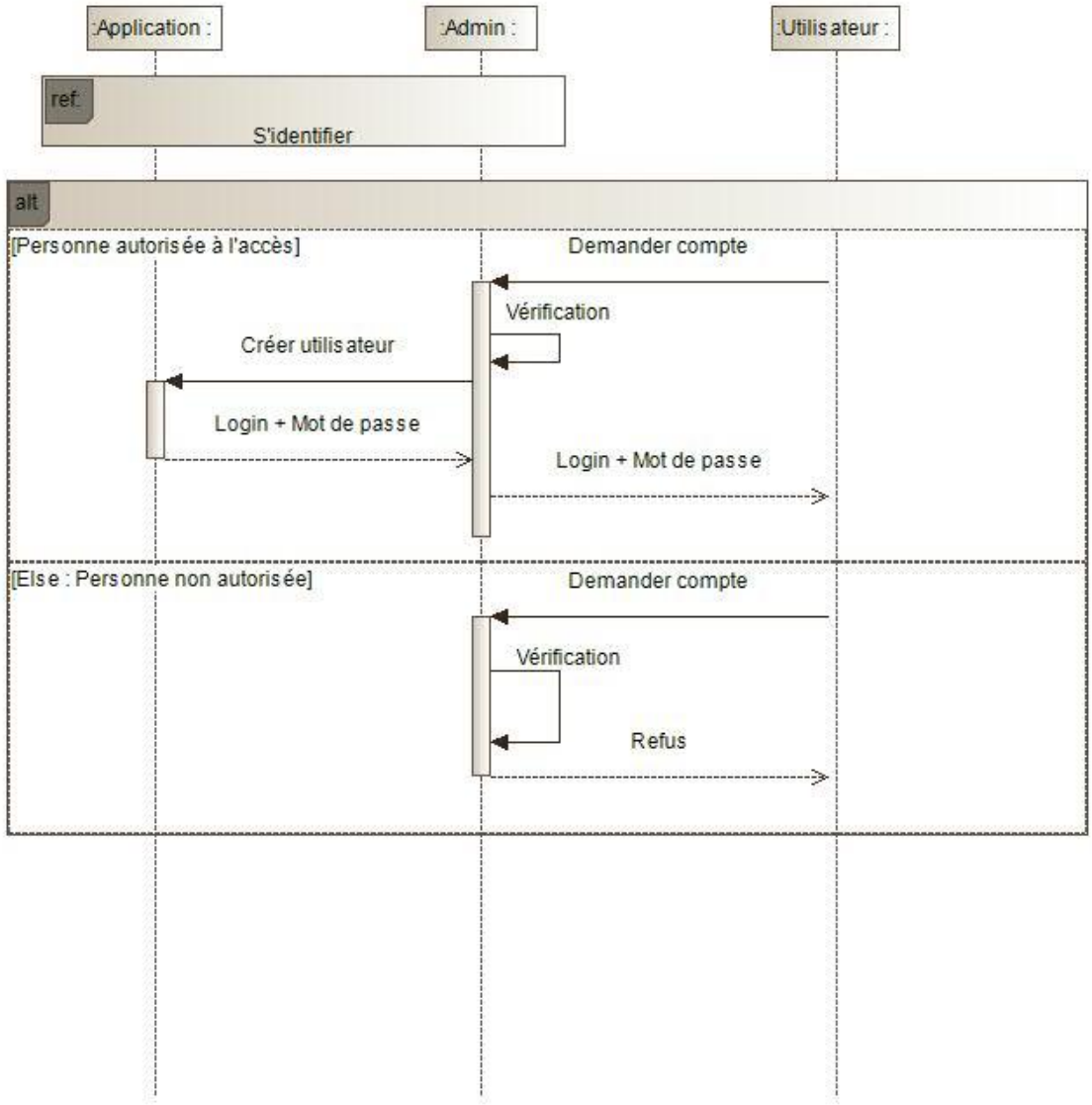


Figure III. 5 : Diagramme de séquence de la demande d'un compte

c) L'usage de l'application

Le patient demande un compte, puis se connecte, il peut changer ses informations personnelles (s'il en a envie) puis dépose son dossier médical, le médecin récupère le dossier déposé par le patient, il peut changer son état (de « positif » a « vacciné » par exemple), puis vient le tour de l'analyste et le chercheur qui ont pour but de récupérer les informations auprès de l'application.

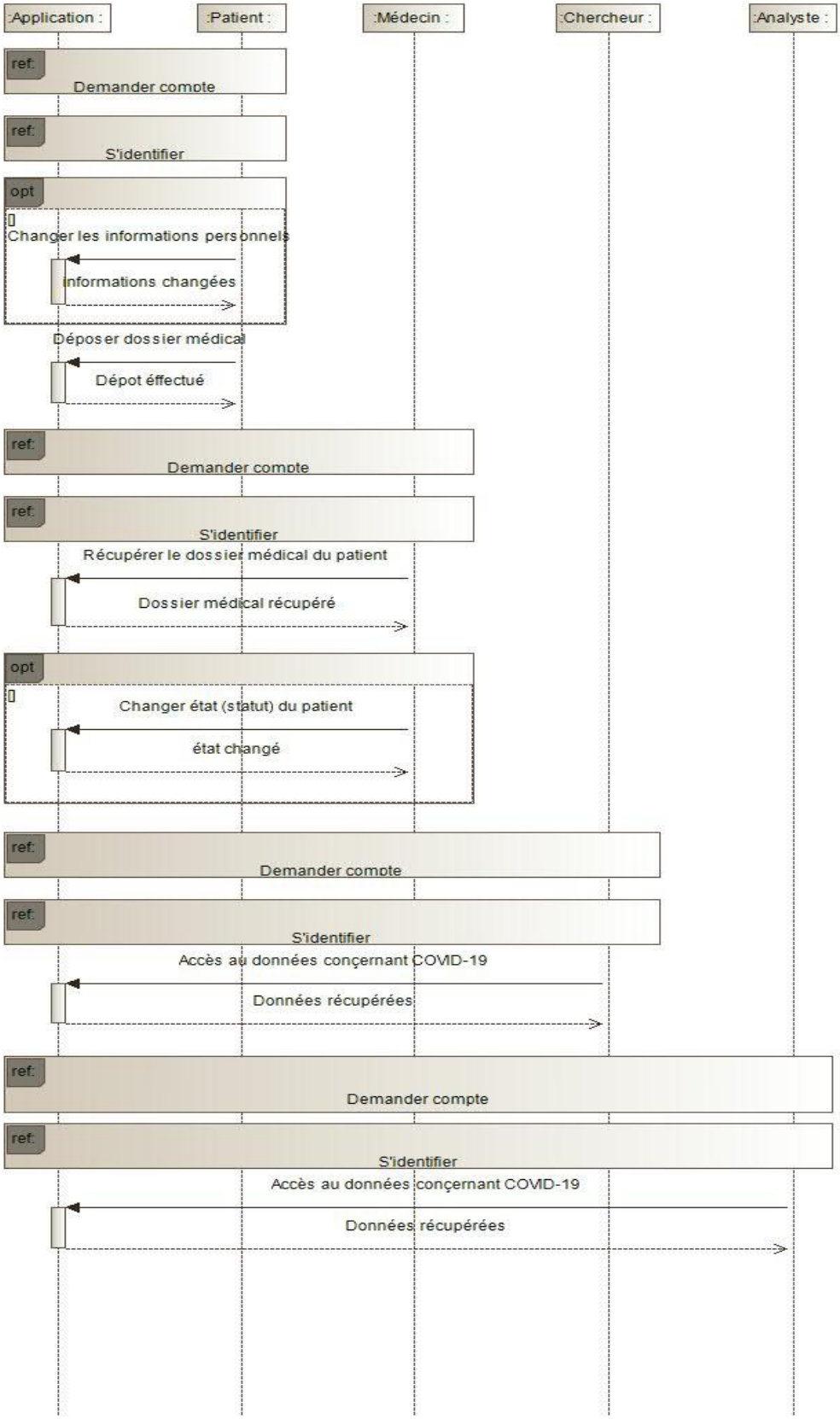


Figure III.6 : Diagramme de séquence concernant l'usage de l'application

III.1.3. Diagramme de classe

À partir des diagrammes précédents, on déduit le diagramme de classe suivant, qui nous permet d'établir notre base de données :

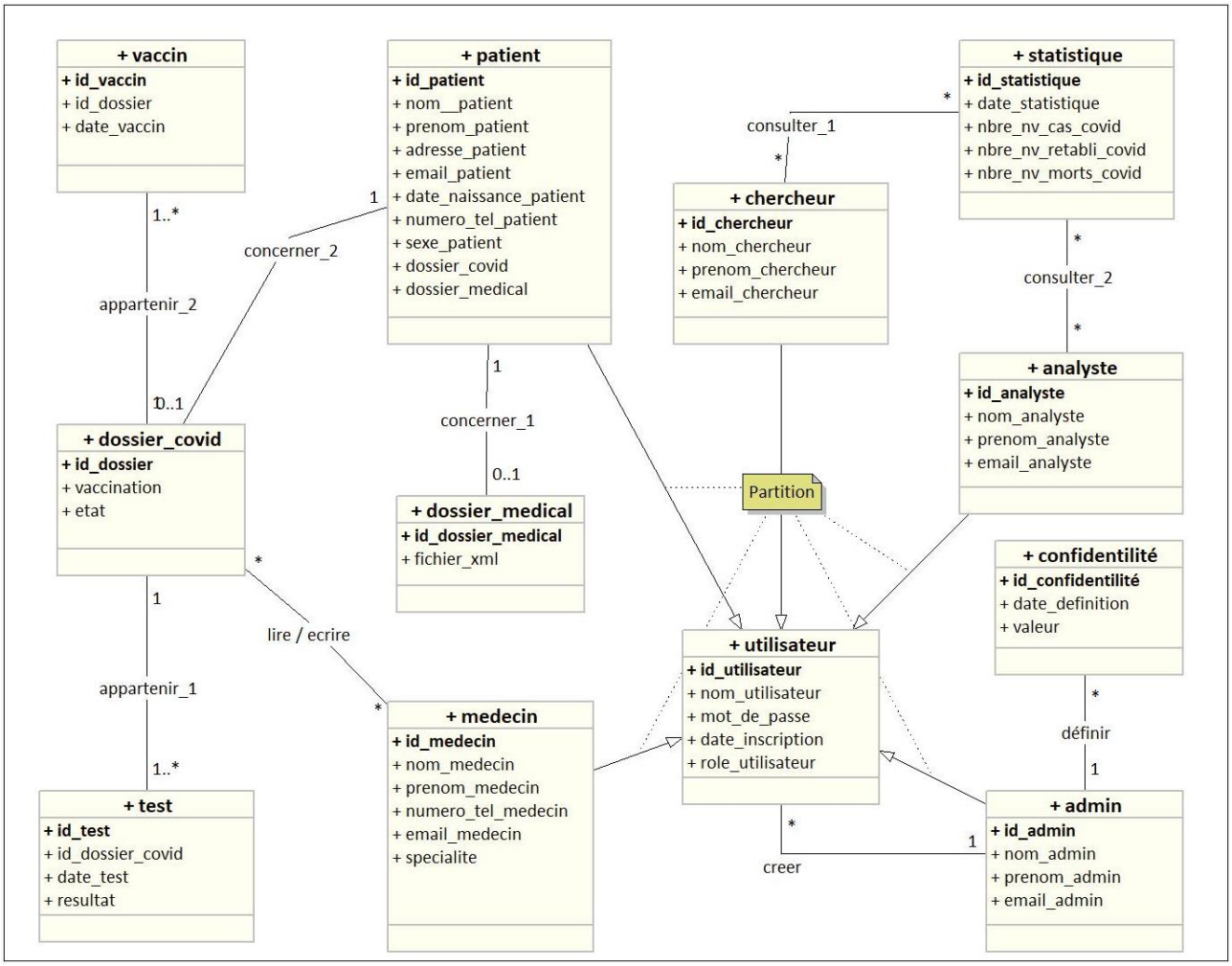


Figure III.7 : Diagramme de classe

III.2. Implémentation

Dans cette partie nous allons citer tous les logiciels qu'on a utilisé pour développer notre application et aussi les outils qui nous ont permis de rédiger ce mémoire, ainsi que le mode de fonctionnement de notre application et toutes les interfaces qui s'y trouvent.

III.2.1. Outils utilisés

Comme toute application il est nécessaire d'utiliser des langages de programmation, des bibliothèques et des logiciels, afin de réaliser notre projet de fin d'études.

a) Logiciels

Ci-dessous se retrouvent les logiciels qu'on a utilisé dans notre projet :

PyCharm :

PyCharm est un environnement de développement intégré pour la programmation en Python. Il permet l'analyse de code et inclut un débogueur graphique. Il permet également la gestion des tests unitaires, l'intégration de logiciels de gestion de versions

Visual Studio Code :

Visual Studio Code, également communément appelé VS Code, est un éditeur de code source créé par Microsoft pour Windows, Linux et MacOs. Les fonctionnalités incluent la prise en charge du débogage, la coloration syntaxique, la complétion intelligente du code, les extraits de code, la refactorisation du code et Git intégré.

Cet éditeur nous a été utile pour créer et éditer le code source de notre application.

Qt Designer :

Qt Designer est l'outil Qt pour concevoir et créer des interfaces utilisateur graphiques (GUI) avec Qt Widgets. On peut composer et personnaliser nos fenêtres ou boîtes de dialogue de manière WYSIWYG (ce que vous voyez, c'est ce que vous obtenez) et les tester à l'aide de différents styles et résolutions.

Ça a servi à concevoir et créer les différentes interfaces graphiques de notre application.

XAMPP :

XAMPP est un ensemble de logiciels permettant de mettre en place un serveur Web local, un serveur FTP et un serveur de messagerie électronique. C'est une distribution freeware qui offre une bonne souplesse d'utilisation et qui est connue pour son installation facile et rapide.

XAMPP a été utile en termes de manipulation de la base de données MySQL de notre application.

Excel :

Excel est un tableur de la suite de productivité Microsoft Office développée et distribuée par Microsoft. La dernière version est Excel 2019. Il est conçu pour fonctionner sur les plates-formes Microsoft Windows, Mac OS X, Android ou Linux.

On s'est appuyé sur Excel pour la création des fichiers CSV / XML.

Looping MCD :

Looping est un logiciel qui permet une implémentation intuitive et très rapide des Modèles Conceptuels de Données (MCD) et la génération automatique en temps réel des Modèles Logiques (MLD) et des requêtes SQL pour créer les tables de base de données correspondantes.

Modelio :

Modelio est un outil UML open source développé par Modeliosoft, basé à Paris, France. Il prend en charge les normes UML2 et BPMN.

Modelio a servi pour la réalisation des diagrammes de cas d'utilisation et de séquence.

b) Langages

On a utilisé principalement deux langages :

Python 3 :

Python est un langage de programmation interprété, multi-paradigme et multi-plateforme. Il favorise la programmation impérative structurée, fonctionnelle et orientée objet.

C'est le langage qu'on a utilisé pour coder notre application.

MySQL 8 :

MySQL est un système de gestion de BDD (Base De Données) relationnelles. Il est distribué avec une double licence GPL et propriétaire.

Ce langage a été utile pour la manipulation de la base de données de notre application.

c) Bibliothèques

Les bibliothèques suivantes nous ont été utiles :

PipelineDP :

PipelineDP est un framework open source Python permettant d'appliquer des agrégations privées différentielles à de grands ensembles de données à l'aide de systèmes de traitement par lots tels qu'Apache Spark, Apache Beam, etc.

Ce framework a été primordial pour appliquer une confidentialité différentielle à notre application.

PyQt 5 :

PyQt est un module libre qui permet de lier le langage Python avec la bibliothèque Qt distribué sous deux licences : une commerciale et la GNU GPL. Il permet ainsi de créer des interfaces graphiques en Python.

Une extension de Qt Creator permet de générer le code Python d'interfaces graphiques.

Ça a servi à concevoir et créer les différentes interfaces graphiques de notre application.

Peewee :

Peewee est un ORM (Object Relational Mapping), qui est un programme informatique qui agit comme une interface entre une application et une base de données relationnelle pour simuler une base de données orientée objet.

On a utilisé Peewee en ce qui concerne la manipulation de la base de données MySQL de notre application.

III.2.2. Présentation de l'application

L'application que nous avons conçue, est structurée sous forme de plusieurs interfaces, chaque type d'utilisateur à accès à l'interface qui lui correspond.

a) Interface de connexion

Grace à cette interface l'utilisateur peut accéder à notre application à partir de son identifiant et son mot de passe

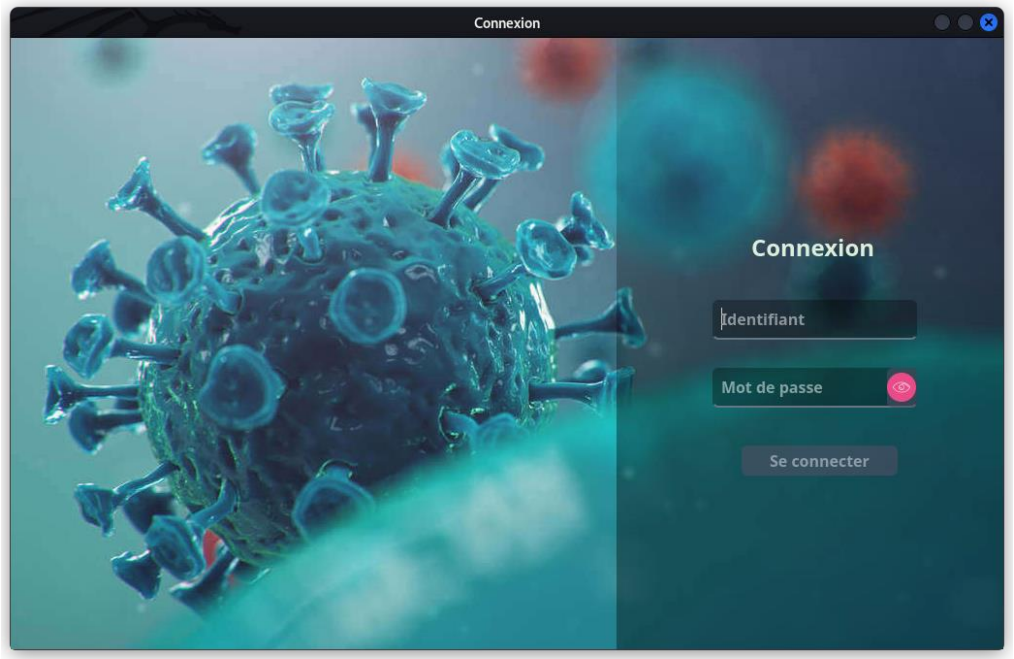


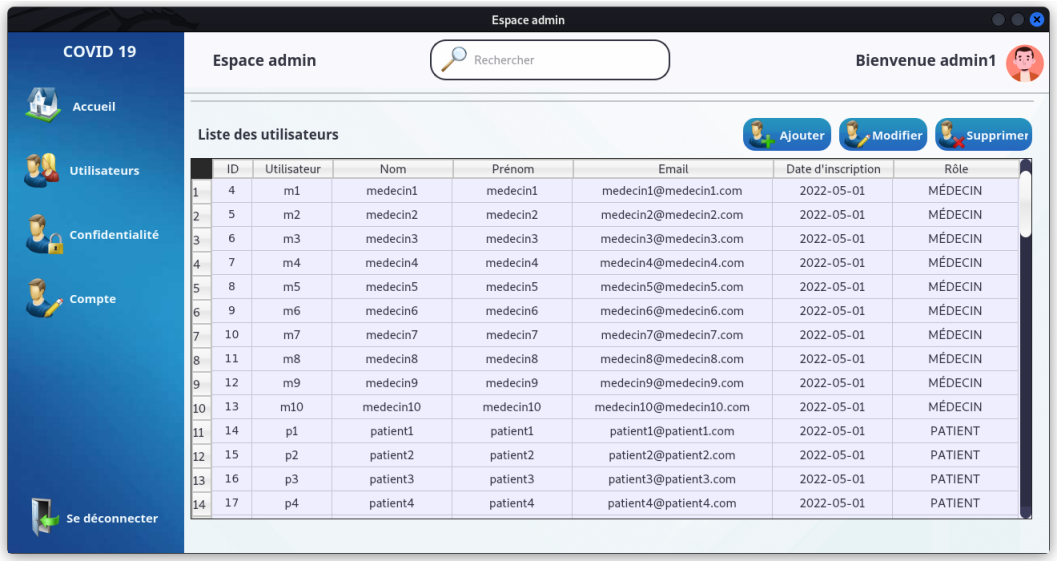
Figure III. 8 : Interface de Connexion

b) Espace administration

Grace à cette interface l'admin peut effectuer l'ensemble des tâches administratives, y compris le niveau de protection de la vie privée pour les patients inscrits.

Figure III. 9 : Espace d'administration

La page d'accueil principale (Onglet Accueil)



	ID	Utilisateur	Nom	Prénom	Email	Date d'inscription	Rôle
1	4	m1	medecin1	medecin1	medecin1@medecin1.com	2022-05-01	MÉDECIN
2	5	m2	medecin2	medecin2	medecin2@medecin2.com	2022-05-01	MÉDECIN
3	6	m3	medecin3	medecin3	medecin3@medecin3.com	2022-05-01	MÉDECIN
4	7	m4	medecin4	medecin4	medecin4@medecin4.com	2022-05-01	MÉDECIN
5	8	m5	medecin5	medecin5	medecin5@medecin5.com	2022-05-01	MÉDECIN
6	9	m6	medecin6	medecin6	medecin6@medecin6.com	2022-05-01	MÉDECIN
7	10	m7	medecin7	medecin7	medecin7@medecin7.com	2022-05-01	MÉDECIN
8	11	m8	medecin8	medecin8	medecin8@medecin8.com	2022-05-01	MÉDECIN
9	12	m9	medecin9	medecin9	medecin9@medecin9.com	2022-05-01	MÉDECIN
10	13	m10	medecin10	medecin10	medecin10@medecin10.com	2022-05-01	MÉDECIN
11	14	p1	patient1	patient1	patient1@patient1.com	2022-05-01	PATIENT
12	15	p2	patient2	patient2	patient2@patient2.com	2022-05-01	PATIENT
13	16	p3	patient3	patient3	patient3@patient3.com	2022-05-01	PATIENT
14	17	p4	patient4	patient4	patient4@patient4.com	2022-05-01	PATIENT

Figure III. 10 : La liste des utilisateurs sous le contrôle administratif (Onglet Utilisateurs)

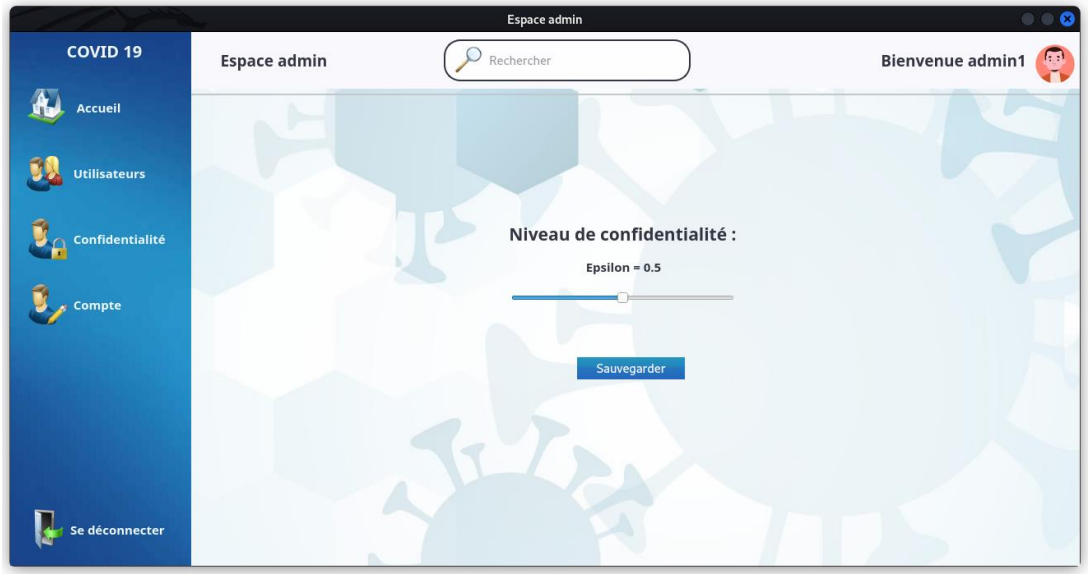


Figure III. 11 : Contrôle de niveau de Confidentialité

Contrôle de niveau de confidentialité qui se traduira en valeur pour Epsilon (Onglet Confidentialité)

c) Espace patient

Grace à cette interface le patient peut s'informer sur le COVID-19, voir ses informations concernant son dossier COVID-19, et aussi voir et modifier ses informations personnelles, ainsi que son dossier médical.

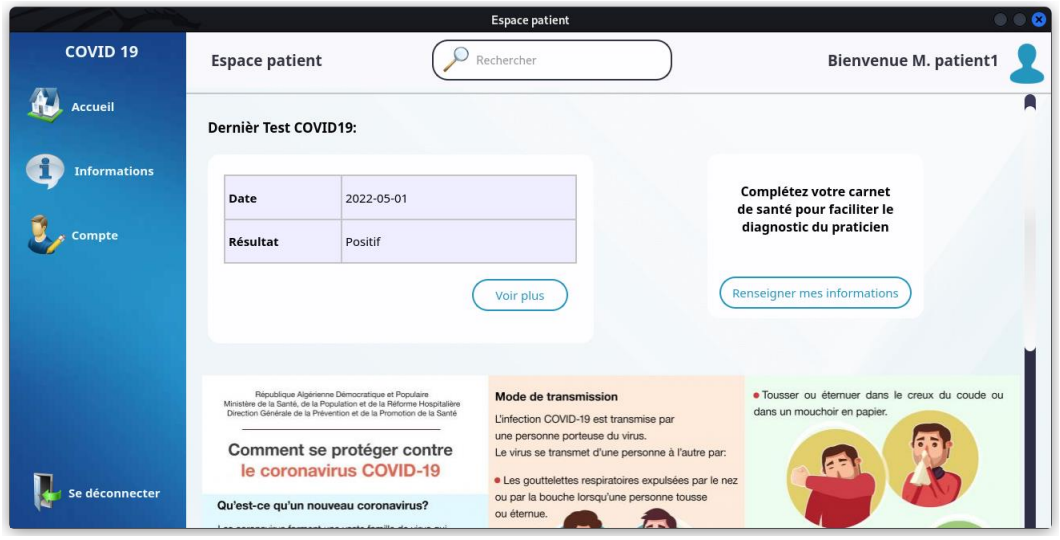


Figure III. 12 : Espace Patient (1)

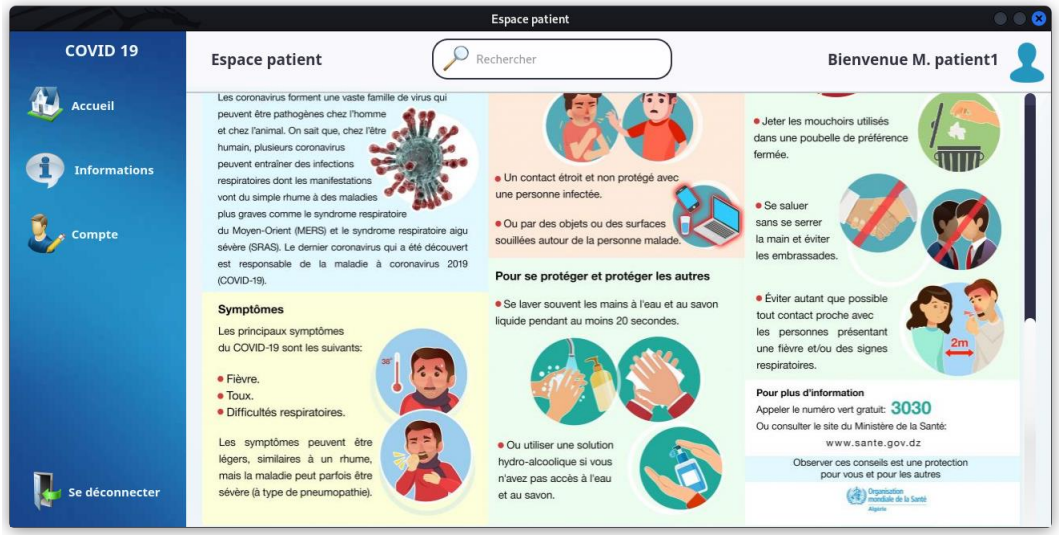


Figure III. 13 : Espace patient (2)

La page d'accueil principale (Onglet Accueil)

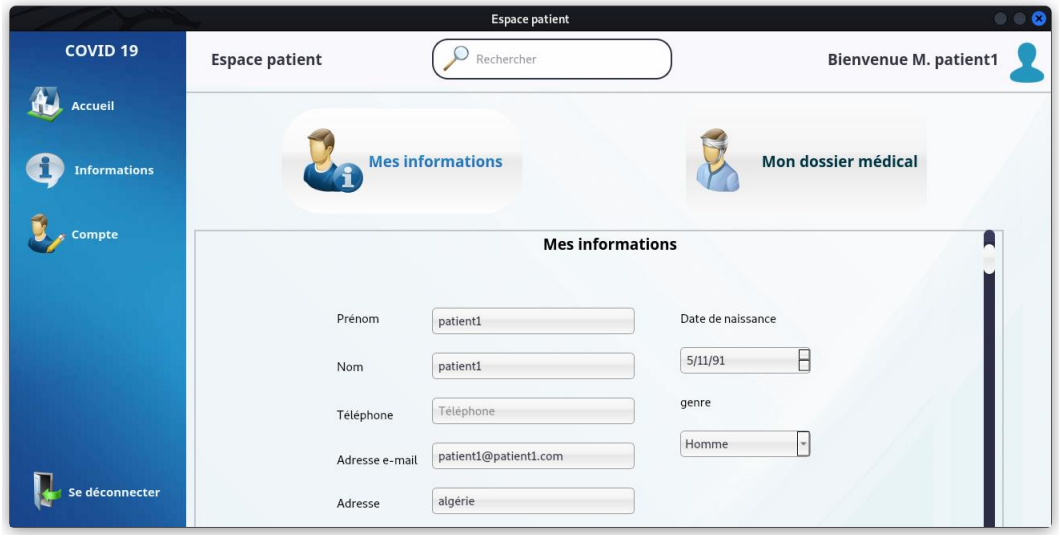


Figure III. 14 : Contrôle des données personnelles

Contrôle de données personnelles (Onglet Compte -> Mes informations)

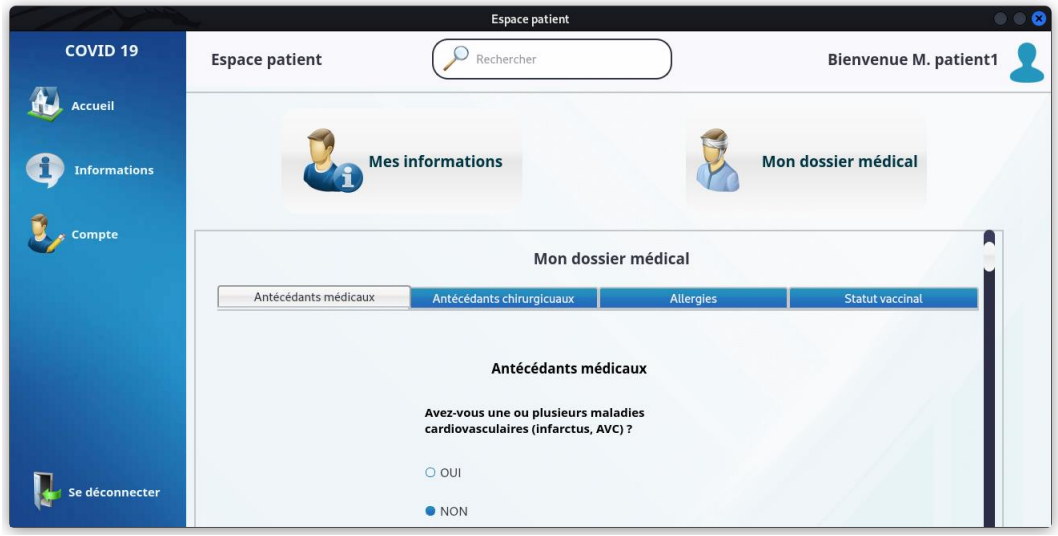


Figure III. 15 : Contrôle des données médicales

Contrôle de données médicales (Onglet Compte -> Mon dossier médical)

d) Espace médecin

Grace à cette interface, le médecin peut s'informer sur les statistiques de COVID-19, voir la liste des patients, ainsi que les informations de chaque patient concernant son dossier COVID-19 et son dossier médical.

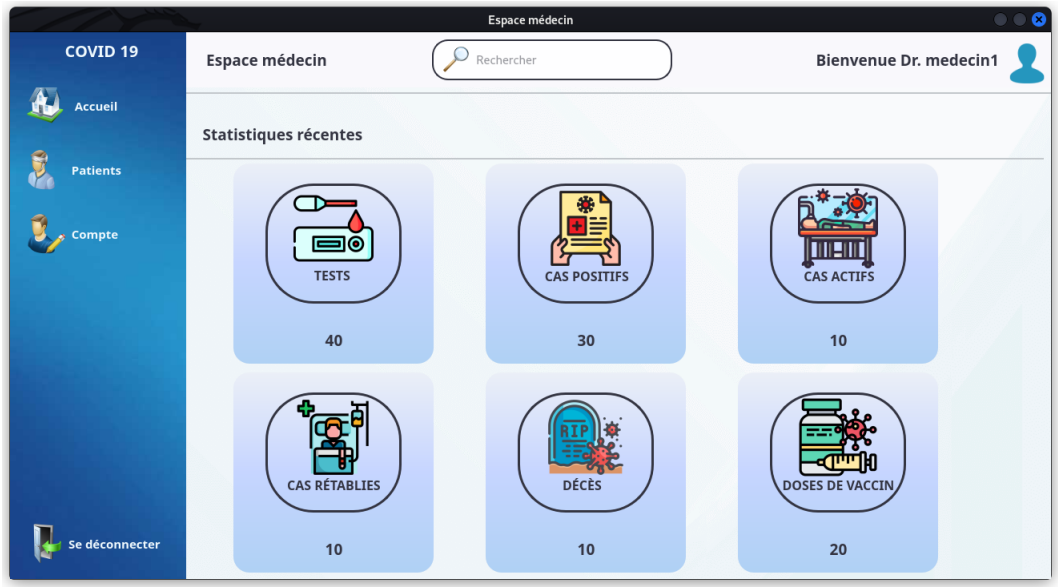


Figure III. 16 : Page des statistiques

La page d'accueil principale avec des statistiques réelles (Onglet Accueil)

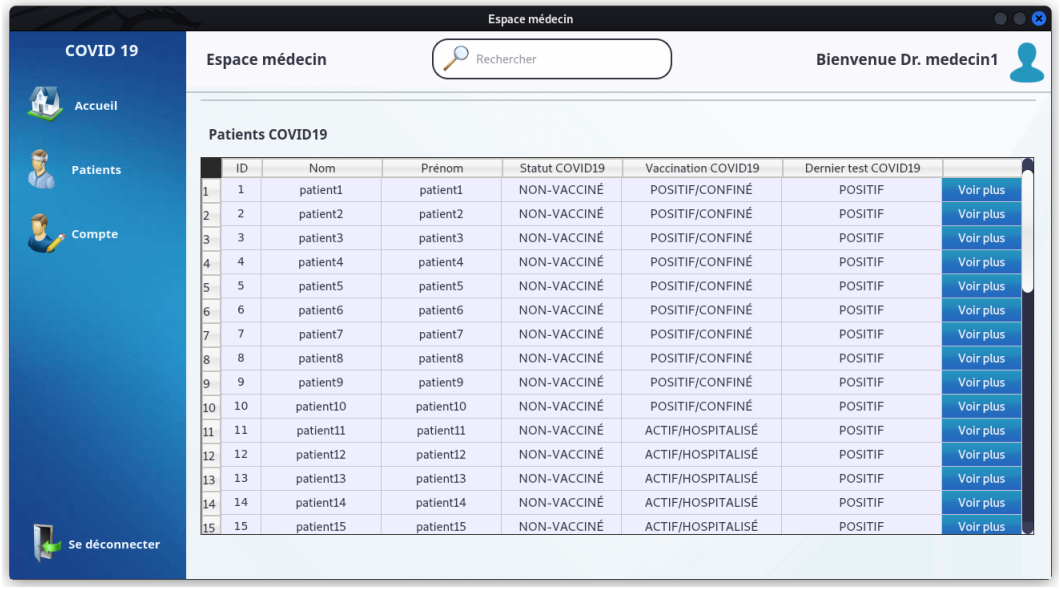


Figure III. 17 : Liste des patients

La liste des patients ainsi que les informations de chaque patient (Onglet Patients)

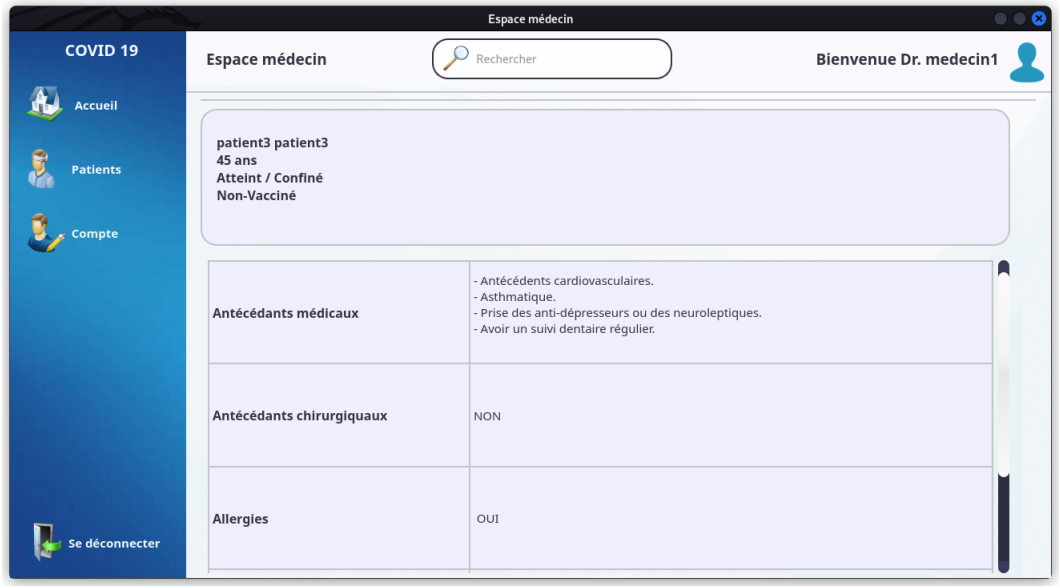


Figure III. 18 : Information sur les patients

Informations d'un patient à propos du COVID-19 ainsi que son dossier médical

(Onglet Patients -> Voir plus)

e) Espace analyste et chercheur

Grace à cette interface, l'analyste et le chercheur peuvent voir l'ensemble des statistiques concernant le COVID-19 ainsi que les informations médicales de chaque patient.

1- Cas sans protection de la vie privée des patients :

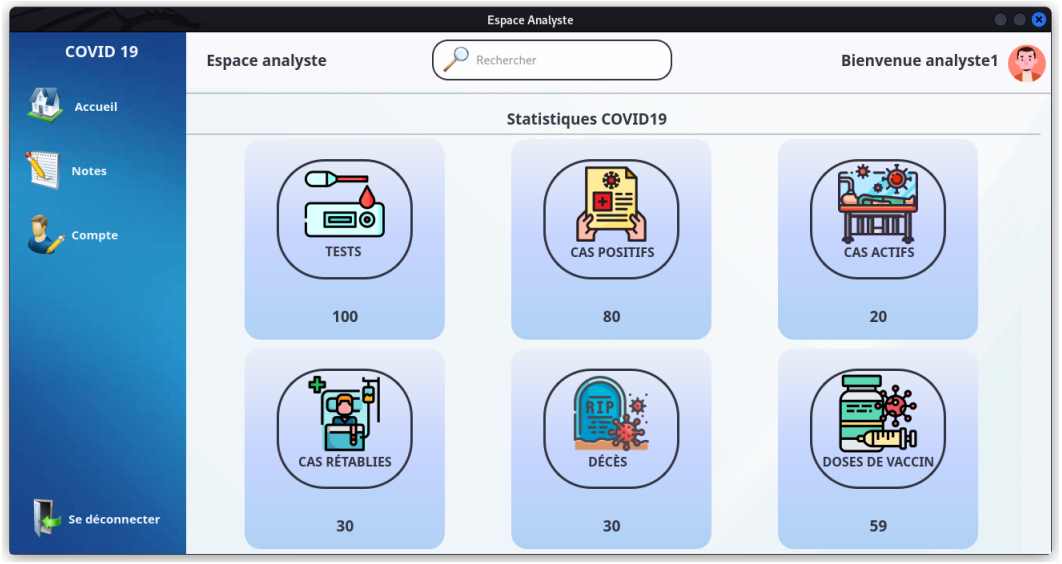


Figure III. 19 : Cas sans protection

La page d'accueil principale (Onglet Accueil)

2- Cas avec protection de la vie privée des patients :

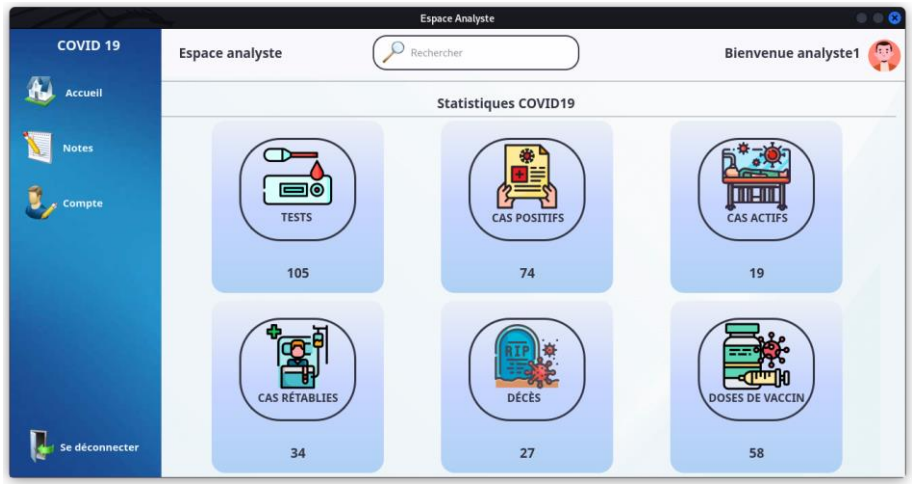


Figure III. 20 : Cas avec protection

Données bruitées (Introduire la confidentialité différentielle)

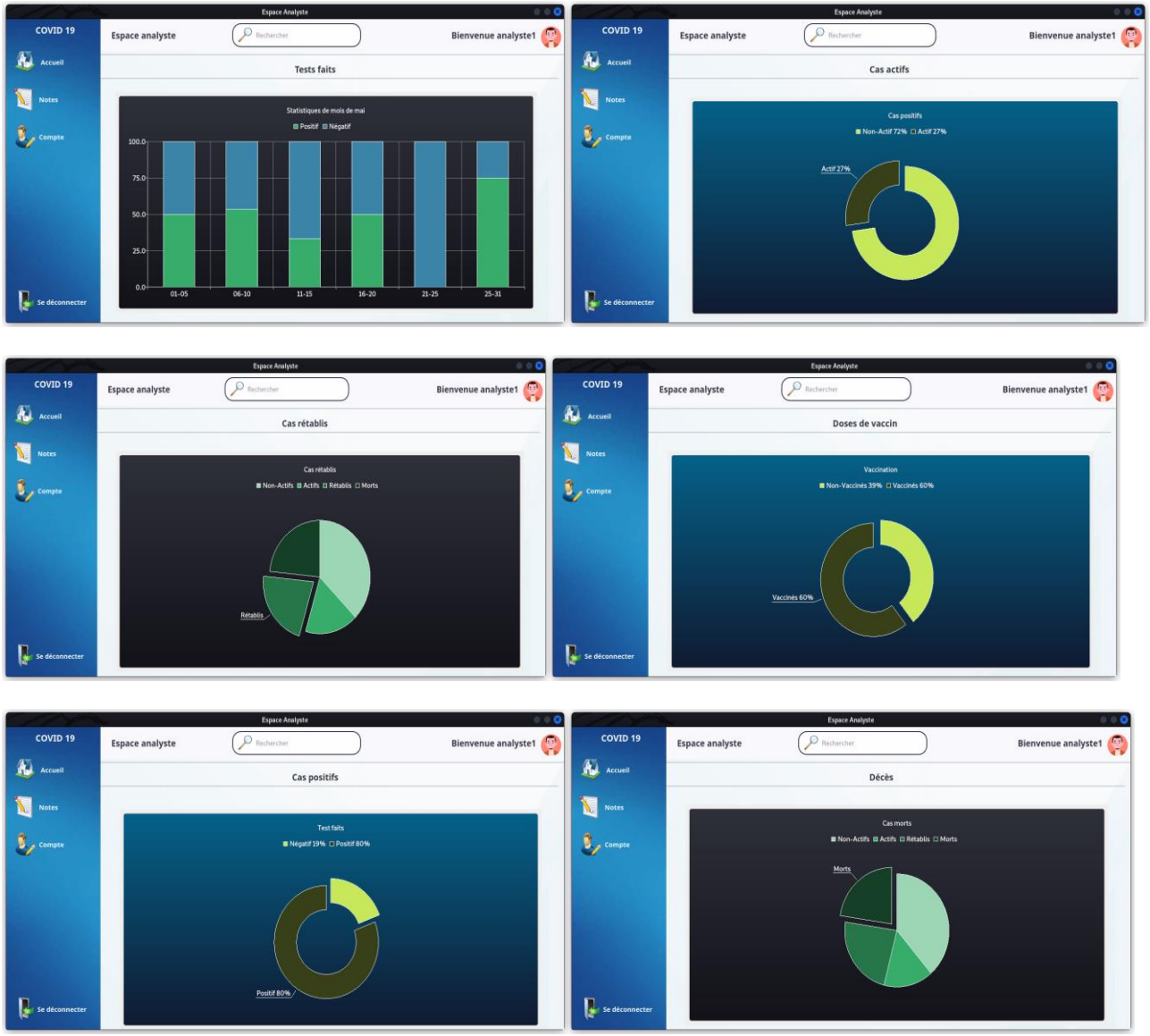


Figure III. 21 : Affichage de graphes avec des données bruitées

III.3. Evaluation des performances

Dans cette partie nous allons comparer entre les chiffres réels (qui se trouvent dans la base de données) et les chiffres affichés à l'analyste et au chercheur suivant les valeurs d'Epsilon, afin d'évaluer la performance de notre application en ce qui concerne la confidentialité différentielle.

III.3.1. Comparaison entre les valeurs exactes et les valeurs bruitées

L'administrateur peut changer la valeur d'Epsilon à sa guise afin que l'analyste et le chercheur ne puissent découvrir les valeurs exactes qui se trouvent dans la base de données de l'application,

Le tableau suivant montre le changement du nombre de cas pour chaque catégorie en fonction de la valeur d'Epsilon.

On tient à préciser que les valeurs inclus dans le tableau et dans les diagrammes, sont valables pour l'analyste et le chercheur.

Epsilon / Cas	0	0.25	0.5	0.75	1
Tests	1000	956	1025	1064	1005
Cas positifs	506	458	532	585	500
Cas actifs	245	245	249	268	266
Cas rétablis	256	276	258	237	259
Cas morts	266	266	212	269	279
Doses de vaccin	505	530	473	502	507

Tableau III. 1 :Changement de données selon la valeur d'epsilon

On convertit le tableau en diagramme à barres, et cela pour chaque valeur d'Epsilon.

Sans confidentialité différentielle : valeurs exactes (sans bruit).

Avec confidentialité différentielle : valeurs approximatifs (avec bruit).

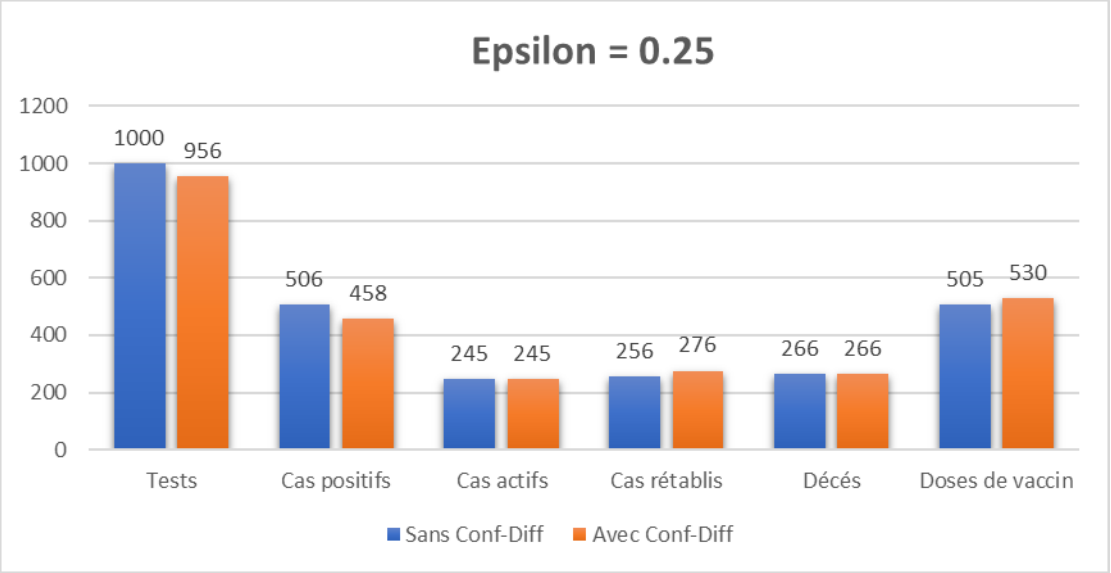


Figure III. 22 : Comparaison entre les données exactes et ceux dont Espilon = 0.25

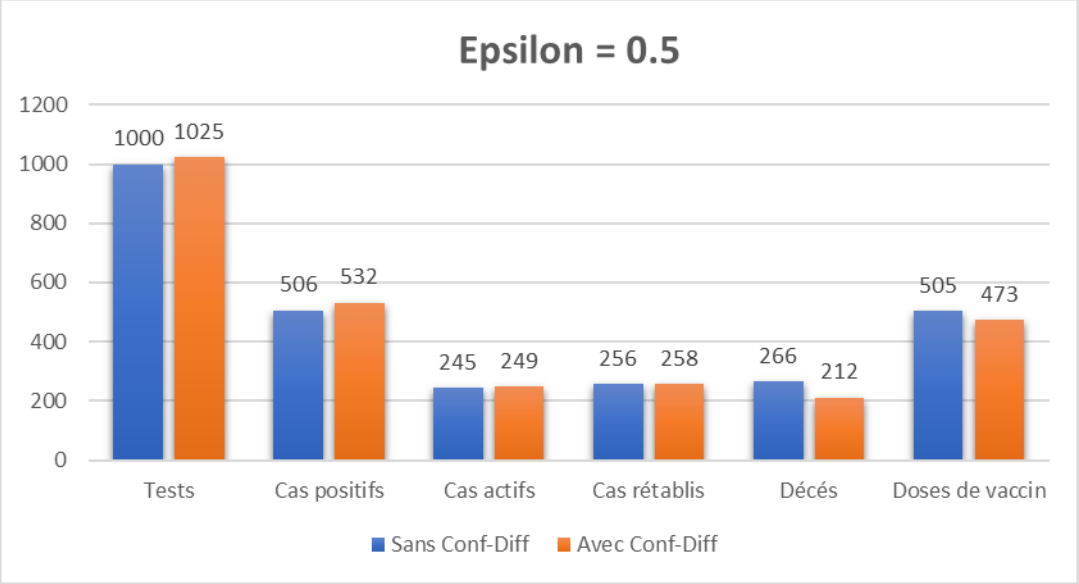


Figure III. 23 : Comparaison entre les données exactes avec celles dont Espilon = 0.5

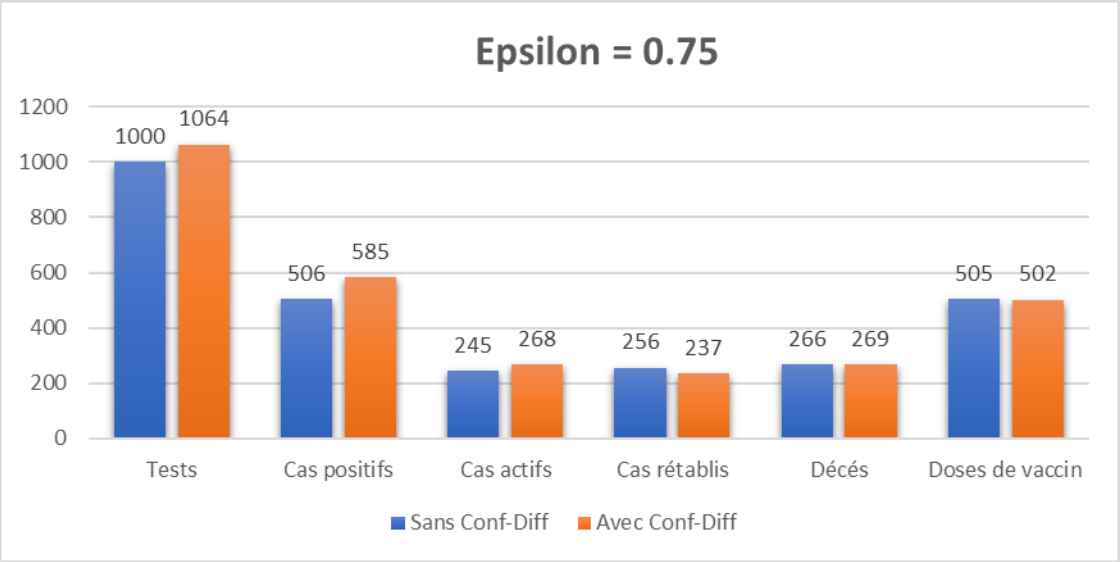


Figure III. 24 : Comparaison entre les données exactes et celles dont Espilon = 0.75

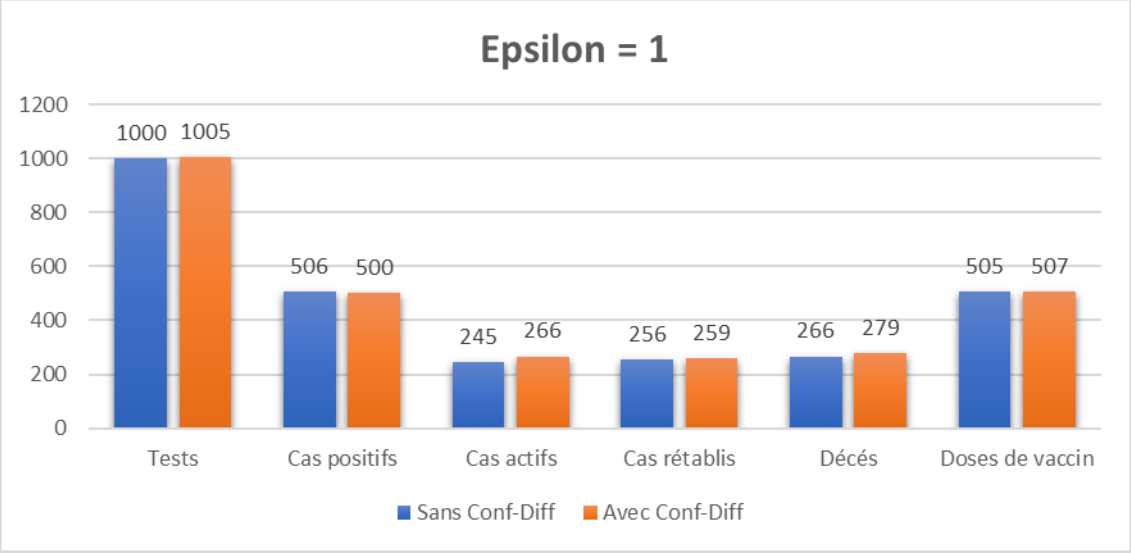


Figure III. 25 : Comparaison entre les données exactes et ceux dont Espilon = 1

On remarque ici que plus la valeur d'Epsilon est petite, plus les chiffres sont crédibles mais la protection des données sensibles diminue.

Dans le cas opposé, plus l'Epsilon est grand, plus la protection des données sensibles est assurée, tandis que la crédibilité des chiffres diminue.

On déduit qu'Epsilon c'est le degré de différence par rapport aux valeurs exactes de la base de données.

Une valeur d'Epsilon qui se rapproche de 0.5 est plus convenable puisque ça permet d'équilibrer entre la véracité des chiffres, et la protection de la sensibilité des données.

Conclusion générale et perspectives

La confidentialité différentielle est un concept très vaste et de nature mathématique, mais le fait de l'avoir considéré comme sujet de recherche, nous a enrichi et formé, cela nous permis de nous familiariser avec le langage python en premier lieu que nous n'avons pas eu le temps de l'apprendre durant les semestres précédents, mais aussi avec la bibliothèque pipelineDP qui est la base du concept du « bruit » de la confidentialité différentielle, et que sans cette bibliothèque notre recherche n'aurait pas été aboutie.

Le fait d'avoir développé une application a été très fructueux pour nous, cela nous a donné l'occasion de nous perfectionner dans les langages python, MySQL et XML et le langage de modélisation UML, comme le dit le proverbe « C'est en forgeant qu'on devient forgeron ».

On a aussi appris des méthodes de sécurité comme RBAC, qui a pour concept d'attribuer a chaque personne le rôle qui lui convient, et chaque rôle correspond à des privilèges pour cette personne, ce qui peut restreindre les risques des erreurs involontaires causées par les utilisateurs de l'application.

Mais pour conclure, il existe toujours des améliorations à envisager pour rendre une application encore performante par exemple :

- Réaliser une application mobile consacrée à l'application afin de pouvoir la rendre plus accessible.
- Introduire la discussion entre le médecin et le patient.
- Ajout du service de paiement en ligne pour l'analyste et le chercheur, les données collectés ne seront plus gratuites.
- Ajout des services techniques de sécurité Comme (Serveur Radius pour l'authentification).

Bibliographie

[1] Bordes J., Bressaud X., Kok Heang K., « Utilisation des données et de l'intelligence artificielle aux Etats-Unis pour lutter contre la maladie »

[2] Belkhamsa A. Yahiaoui M., « La protection de la vie privée dans le Big Data », Université AMO de Bouira, 2019/2020.

[3] Nguyen B., « Techniques d'anonymisation », pp.43-50, Décembre 2014.

[4] Rioux J., « Un modèle rétroactif de réconciliation utilité-confidentialité sur les données d'assurance », Université de Montréal, Avril 2016.

[5] Feten Ben Fredj. Méthode et outil d'anonymisation des données sensibles. Cryptographie et sécurité [cs.CR]. Conservatoire national des arts et métiers – CNAM ; Université de Sfax (Tunisie). Faculté des Sciences économiques et de gestion, 2017.

[6] Charest A., « Protection statistique de la confidentialité des données », Université Laval, 29 octobre 2015.

[7] Tarek Benkhelif. Publication de données individuelles respectueuse de la vie privée : une démarche fondée sur le co-clustering. Cryptographie et sécurité [cs.CR]. Université de Nantes, 2018. Français.

[8] Solenn Brunet. Conception de mécanismes d'accréditations anonymes et d'anonymisation de données. Cryptographie et sécurité [cs.CR]. Université de Rennes 1 [UR1], 2017. Français.

[9] Hajaralsadat Torabian, « Protecting Sensitive Data using Differential Privacy and Role-based Access Control », Université Laval (Québec, Canada), 2016.

Webographie

[10]. Dépôt universitaire de mémoires après soutenance : lien https://dumas.ccsd.cnrs.fr/dumas-02945116#:~:text=La%20e%2Dsant%C3%A9%20englobe%20l,%C3%A0%20la%20gu%C3%A9rison%20des%20patients

Dernière visite : 29 avril 22

[11]. SYNOX : lien https://www.synox.io/votre-secteur/sante-connectee/

Dernière visite : 29 avril 22

[12]. L'organisation mondiale de la santé : lien https://www.who.int/fr/health-topics/coronavirus/coronavirus#tab=tab_1

Dernière visite : 7 mars 22

[13]. l'organisation mondiale de la santé : lien https://www.who.int/fr/news-room/feature-stories/detail/new-who-technical-package-to-help-countries-improve-health-data-for-covid-19-response-and-beyond#:~:text=%C2%AB%20L'outil%20technique%20SCORE%20fournit,for%20Health%20pour%20fournir%20cette Dernière visite : 7 mars 22

[14]. Google news :lien https://news.google.com/covid19/map?hl=fr&mid=%2Fm%2F02j71&gl=FR&ceid=FR%3AfrDernière visite : 7 mars 22

[15]. Our World In Data : lien https://ourworldindata.org/coronavirus Dernière visite : 7 mars 22

[16]. Google : lien https://www.google.com/search?q=le+nombre+de+cas+du+covid-19+au+monde&sxsrf=APq-WBv_wMFyflZh3tDw9cH-3-cvjV676A%3A1647295556422&ei=RLwvYt-tGcP9kwXeoILQCg&ved=0ahUKEwjf3-uTzsb2AhXD_qQKHV6QAKoQ4dUDCA4&uact=5&oq=le+nombre+de+cas+du+covid-19+au+monde&gs_lcp=Cgdnd3Mtd2l6EAM6BwgAEEcQsAM6BggAEAcQHjoGCAAQCBAeOgYIIxAnEBM6CAgAEAgQBxAeOggIABAHEAUQHkoECEEYAEoECEYYAFCQCli8aWDUdGgDcAF4AYAB8wOIAcUlkgEKMC4xMy4zLjQuMZgBAKABAcgBB8ABAQ&sclient=gws-wiz Dernière visite : 7 mars 22

[17]. PIABLOG : lien https://www.privateinternetaccess.com/blog/our-seven-privacies-the-many-important-facets-of-privacy/?fbclid=IwAR0_5vUlyl_GCovnTN9H8njS-KthU7TLrAkzcUiFfiIKUxAcphNZDXh4qe0#:~:text=There%20are%20seven%20distinct%20important,invade%20it%20without%20your%20consent Dernière visite : 30 avril 22

[19]. Murielle Cahen : lien https://www.murielle-cahen.com/publications/p_vieprivee.asp

Dernière visite : 3 mai 22

[20]. Statistique Canada : lien https://www.statcan.gc.ca/fr/science-donnees/reseau/protection-vie-privee Dernière visite : 15 mars 22

[21] : Data Value Consulting : lien https://datavalue-consulting.com/ethique-ia/

Dernière visite : 4 mai 22

[22] : JAFWIN Data : lien https://jafwin.com/2020/04/03/lessentiel-a-savoir-sur-la-confidentialite-differentielle/#:~:text=traite%20les%20donn%C3%A9es.-,La%20confidentialit%C3%A9%20diff%C3%A9rentielle%20globale,avec%20un%20algorithme%20diff%C3%A9rentiellement%20priv%C3%A9 Dernière visite : 4 mai 22

[23] : The astrology page : lien https://fr.theastrologypage.com/role-based-access-control

Dernière visite : 4 mai 22

[24] : IONOS : lien https://www.ionos.fr/digitalguide/serveur/securite/quest-ce-que-le-role-based-access-control-rbac/#:~:text=Le%20fonctionnement%20du%20Role%20Based%20Access%20Control,-Avant%20de%20pouvoir&text=Cela%20inclut%20l'%C3%A9tablissement%20pr%C3%A9cis,Autorisations%20dans%20les%20applications

Dernière visite : 4 mai 22