Mémoire de recherche pour l'obtention d'un master MIAGE parcours Système d'Information et Innovation (S2I)

Université Paris 1 Panthéon Sorbonne

Les aspects de vie privée dans l'échange de données médicales avec la Blockchain

ATTAL Ruben
Tuteur enseignant : Herbaut Nicolas
Maître d'apprentissage : SALOME Christophe
Année : 2021/2022

ATTESTATION SUR L'HONNEUR DE NON PLAGIAT

Je soussigné(e) ATTAL Ruben déclare sur l'honneur que ce mémoire est le fruit d'un travail personnel et que je n'ai ni contrefait, ni falsifié, ni copié tout ou partie de l'œuvre d'autrui afin de la faire passer pour mienne.

Toutes les sources d'information utilisées (supports papiers, audiovisuels et numériques) et les citations d'auteur ont été mentionnées conformément aux usages en vigueur. Je suis conscient(e) que le fait de ne pas citer une source ou de ne pas la citer clairement et complètement est constitutif de plagiat, que le plagiat est considéré comme une faute grave au sein de l'Université et qu'il peut être sévèrement sanctionné.

Date et signature :

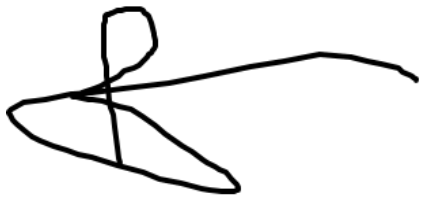


Table des matières

Remerciements

Je tenais à remercier toutes les personnes qui m'ont aidé et contribué pour la rédaction de ce mémoire.

Je souhaite pour commencer remercier sincèrement M. Nicolas Herbaut, mon tuteur enseignant, maître de conférences, responsable du M1 MIAGE Sorbonne en Apprentissage, enseignant à Paris 1 Panthéon Sorbonne et à l'UFR de Mathématique et Informatique, pour m'avoir accordé sa disponibilité, son partage de connaissance et ses précieux conseils qui m'ont permis d'avancer grandement tout au long de la réalisation de ce mémoire.

Je remercie également toute l'équipe pédagogique de l'université Paris 1 Panthéon Sorbonne ainsi que le CFA AFIA pour leurs encadrements, les connaissances et les techniques que j'ai pu assimiler pendant ma formation.

Pour finir, je remercie les membres de mon équipe, la direction technique transport au sein d'Axa France où j'ai pu effectuer mon alternance et plus particulièrement Monsieur BOUKALA Amine pour son encadrement durant mon alternance, DUCCAFY Cyril et SALOME Christophe pour leur accompagnement et conseils.

Avant-propos :

La crise sanitaire a été l'occasion de revoir la gestion des systèmes hospitaliers dans le monde. La digitalisation de la santé s'est envolée, et ce, de façon plus intense depuis la crise covid. L'utilisation et le partage de nos données de santé deviennent un sujet majeur prenant une place très importante. La mise en place de dossiers de santé électronique est un sujet de plus en plus actuel avec certaines complexités. Comme nous allons le voir, le traitement de données, mais surtout de données aussi sensibles que le sont les données de santé nécessite de répondre à de nombreux critères en termes de sécurité pour ne pas compromettre la vie privée des patients.

Pour répondre à ces critères, il s'avère que la blockchain est une solution prometteuse et son intégration dans des solutions applicative de mise en place de dossiers de santé électronique, également appelé DSE, commence à voir le jour dans la littérature.
Cependant, il semble que les solutions apportées ne prennent pas aussi sérieusement qu'il le faudrait l'ampleur et le niveau d'exigence que requiert la notion de confidentialité.
Cette technologie étant récente et complexe, les solutions proposées basées sur des standards et architectures reconnus par la communauté seraient plus simples à mettre en place et auraient plus de chance d'être intégrées dans un système de santé.

Ainsi, à travers cette revue systématique de la littérature (SLR) nous étudierons les études primaires discutant de cette problématique afin de voir comment est abordée la notion de confidentialité dans l'échange de données médicales utilisant la blockchain.

Abstract :

The health crisis was an opportunity to review the management of hospital systems around the world. The digitization of health care has taken off worldwide and more intensely since the covid crisis. The use and sharing of our health data is becoming a major topic taking a very important place. The implementation of electronic health records is an increasingly current topic with certain complexities. As we will see, the processing of data, especially data as sensitive as health data, requires meeting many criteria in terms of security so as not to compromise the privacy of patients.

To meet these criteria, blockchain is a promising solution and its integration into application solutions for the implementation of electronic health records is beginning to appear in the literature.
However, it seems that the solutions provided do not take as seriously as they should the scope and level of requirement that the notion of confidentiality requires.
As this technology is recent and complex, the proposed solutions based on standards and architectures recognized by the community would be simpler to implement and would have a better chance of being integrated into a healthcare system.

Thus, through this systematic review of the literature (SLR) we will study the primary studies discussing this issue in order to see how the notion of confidentiality is addressed in the exchange of medical data using blockchain.

I. Introduction

1) Contexte

Les données concernant la santé sont très précieuses pour développer le savoir médical et prolonger la vie des patients. Les nouvelles technologies jouent un rôle essentiel dans le développement des systèmes de santé car elles permettent d'obtenir des données toujours plus précises et de les récupérer via plusieurs supports tels que les équipements médicaux, les objets connectés… Selon une enquête IPSO réalisée en 2016 sur un panel de 2000 Français, 42% déclarent "qu'il est fréquent que les médecins qui les suivent en ville et à l'hôpital ne communiquent pas entre eux, 38% en ce qui concerne les personnels de santé hospitaliers entre eux" [39]. Ce résultat soulève la question de création d'une solution pour pallier à ce besoin. Pour partager ces données, les dossiers de santé électronique, également appelé dossier médical partagé se sont mis en place. Les DSE sont des dossiers de santé mis à jour en temps réel, centrés sur le patient qui rendent les informations disponibles de façon immédiate pour les utilisateurs autorisés.
Le service public français est en train de mettre en place une application nommée monespacesante.fr pour stocker et partager ces documents et données de santé en toute confidentialité. Sauf contre-indication, cette application crée un DSE à toute la population française. Les prestataires de soins de santé utilisent de plus en plus les DSE. Ainsi, ce marché devrait bientôt atteindre 40 milliards de dollars d'ici 2022 [8].

Cependant, cette adoption des DSE expose la vie privée des patients et la sécurité de leurs informations à un risque de violation des données. La cybersanté est un enjeu de plus en plus important et plusieurs solutions pour sécuriser ces informations sont envisagées. La question de la confidentialité et la sécurité pour l'échange de données médicales devient une réelle problématique.

Il est important de prendre conscience que la notion de confidentialité, privacy en anglais, est un terme qui induit plusieurs concepts. Afin de rendre compte de toute cette complexité, le terme sera défini dans la section 2 Protocole de recherche de ce mémoire.

Les données médicales tout comme toutes données à caractères personnelles sont des informations extrêmement sensibles et personnels, leur protection est donc soumise à de nombreuses lois internationales. Le RGPD en France [40], le CCPA en Californie [41], LGPD au Brésil [42] ou encore PIPL en Chine [43], de nombreux pays mettent en place des législations pour régir leurs utilisations et assurer la confidentialité des données de leurs citoyens. Il existe également des réglementations concernant directement le traitement de données de santé comme l'HIPAA (Health Insurance Portability and Accountability Act) qui est une série de normes réglementaires fédérales américaines [44] ou encore la certification HDS en France (Certification Hébergeur de Données de Santé) [45] avec pour objectif de renforcer la protection des données de Santé à caractère personnel et construire un environnement de confiance autour de l'E-santé et du suivi des patients.

2) Définition de la problématique

En raison de ses propriétés uniques qui permettent de concevoir des architectures et des systèmes innovants, la blockchain a cette capacité à fonctionner sans tiers en instaurant la confiance grâce à l'utilisation d'applications décentralisées. De plus, la mise en place du système contrats intelligents ouvre la voie au développement d'un accès DSE efficace avec des méthodes de contrôle pour prendre en charge l'identification, l'authentification et l'autorisation sécurisées des clients [46]. La blockchain pourrait être une technologie innovante pouvant répondre à cette problématique.

Une communauté de chercheurs a commencé à se pencher sur la question en proposant la blockchain pour sécuriser un réseau de partage et d'échange de données médicales.

La confidentialité des données est l'une des principales limites de la blockchain. Toutes les informations sur une blockchain sont disponibles pour les participants au réseau blockchain. Il n'y a pas d'utilisateur privilégié au sein du réseau blockchain, peu importe que la blockchain soit publique, en consortium ou privée. Sur une blockchain publique, les nouveaux participants peuvent rejoindre librement le réseau blockchain et accéder à toutes les informations enregistrées sur la blockchain.

De plus, il semble que dans la littérature existante concernant l'utilisation de la blockchain pour l'échange de données médicales, les problèmes liés à la sécurité et à la confidentialité des données dans les applications de santé sont très brièvement si ce n'est même pas abordé. De plus, s'ils le sont, ils ne sont pas souvent soumis à une réglementation quelconque, mais à une interprétation propre de la confidentialité, selon le chercheur. Les solutions apportées ne sont pas donc créées pour être en accord avec les lois misent en place, mais celles évoquées par le chercheur.

Par l'intermédiaire de la revue systématique de la littérature, ce mémoire a pour but de répondre à la problématique suivante :

"Comment est abordée la notion de confidentialité dans l'échange de données médicales utilisant la blockchain ?"

Cette SLR nous permettra de tester certaines hypothèses qui nous aideront à répondre à cette question de recherche. Ces hypothèses seront formulées dans la section 3 Protocole de recherche.

Afin de répondre à cette problématique, nous procéderons en plusieurs parties. La section 2 donne une définition des éléments clés de cette revue, soit la blockchain, les DSE, la notion de confidentialité et notamment les principales contraintes du RGPD. La section 3 présente le protocole de recherche qui a été adopté pour effectuer cette étude systématique de littérature. La section 4 présente l'extraction et l'analyse des données de cette recherche. Enfin, en section 5 nous discuterons les résultats obtenus pour finir par conclure cette revue dans la section 6.

II. Définition des termes clés

Ce mémoire étant une revue systématique de littérature abordant les notions comme la privacy / confidentialité, la blockchain et les dossiers de santé électronique DSE, nous allons donc définir ces termes clés.

1) La notion de privacy / confidentialité

La confidentialité a été définie par l'Organisation internationale de normalisation (ISO) comme *“le fait de s'assurer que l'information n'est accessible qu'à ceux dont l'accès est autorisé”*.
La confidentialité va de pair avec la vie privée, elle est principalement utilisée en terme juridique pour parler de droit à la confidentialité et protection de la vie privée. Cela correspond aux moyens mis en œuvre pour protéger les données à caractères personnelles des individus. Les données à caractères personnelles sont "toutes les informations relatives à une personne physique susceptible d'être identifiée, directement ou indirectement.” selon la CNIL (Commission nationale de l'informatique et des libertés).

Pour légiférer sur l’utilisation des données à caractères personnelles, beaucoup de réglementations se sont mises en place dans le monde. Sur le territoire de l’Union Européenne par exemple, c'est le RGPD qui encadre le traitement des données personnelles.
Voici la réglementation concernant la protection des données de santé avec le RGPD selon la CNIL [47] :

- Limiter l’accès aux données de santé de vos patients : seules certaines personnes sont autorisées, au regard de leurs missions, à accéder à celles-ci
- Les données que vous collectez sur vos patients doivent être conservées pour une durée déterminée.
- Respecter des règles de sécurité pour protéger les données des patients contre des accès non autorisés (ex : utilisation de la carte professionnelle de santé, mot de passe personnel, utilisation d’un système de chiffrement fort, etc)
    - un engagement cryptographique ;
    - une empreinte de la donnée obtenue par une fonction de hachage à clé ;
    - un chiffré de la donnée
- Vous devez tenir un registre des activités de traitement et le renseigner

Ci-dessous précisions concernant les droits d’accès au DSE [48] :

Précision dans le cadre des données de santé : Il n’y a pas besoin de recueillir le consentement des patients pour collecter et conserver les données de santé les concernant, dans la mesure où leur collecte et leur conservation sont nécessaires aux diagnostics médicaux et à la prise en charge sanitaire ou sociale des patients concernés.

Qui peut demander l'accès au dossier médical ?
L'accès au dossier médical peut être demandé auprès du professionnel de santé ou de l'établissement de santé, par la personne concernée, son ayant droit en cas de décès de cette personne, le titulaire de l'autorité parentale, le tuteur ou le médecin désigné comme intermédiaire.

Quelles sont les informations communicables ?
Toute personne a accès à l'ensemble des informations concernant sa santé,

La question du responsable de traitement se pose également. La CNIL définie le responsable de traitement de la façon suivante :

"*Le responsable de traitement est la personne morale (entreprise, commune, etc.) ou physique qui détermine les finalités et les moyens d'un traitement, c'est-à-dire l'objectif et la façon de le réaliser. En pratique et en général, il s'agit de la personne morale incarnée par son représentant légal.*" [49]

La notion de traitement de données personnelles étant très large, précision pareillement cette notion. La CNIL définie le traitement de données personnelles comme "une opération ou ensemble d'opérations, portant sur des données personnelles, quel que soit le procédé utilisé (collecte, enregistrement, organisation, conservation, adaptation, modification, extraction, consultation, utilisation, communication par transmission, diffusion ". [40]

Le RGPD, et plus généralement le]s principes classiques de la protection des données, ont été conçus dans un monde où la gestion des données est centralisée au sein d'entités déterminées. Le modèle décentralisé de gouvernance des données de la technologie Blockchain et la multiplicité des acteurs intervenant dans le traitement de la donnée complexifient la définition des rôles de chacun.

La CNIL constate toutefois que les participants, qui ont un droit d'écriture sur la chaîne et qui décident de soumettre une donnée à la validation des mineurs peuvent être considérés comme responsables de traitement.

Cela peut poser un vrai problème sur la gestion des responsables de traitement, car la blockchain fonctionne sur des processus tels que le proof of work, proof of stake ou des mineurs valident les transactions...

La CNIL répond à cette question en indiquant que les mineurs, se limitant à valider les transactions soumises par les participants sans intervenir sur le contenu de ces transactions, ne sont pas considérés comme des responsables de traitement.

Veuillez trouver via cette référence [50] le document suivant "La Blockchain : quelles solutions pour un usage responsable en présence de données personnelles ?"

Il existe plusieurs autres réglementations relatives à la protection des données comme le CCPA en Californie, LGPD au Brésil ou encore PIPL en Chine.

La cybersanté prenant une ampleur considérable avec toujours plus de numérisation, les dispositifs pour veiller à la protection des données personnelles sont de plus en plus développés.

Il existe aussi des lois spécifiques à la protection des données de santé comme l'HIPAA (Health Insurance Portability and Accountability Act) établis en 1996 aux Etat-Unis. L'HIPAA est une série de normes réglementaires fédérales américaines qui décrivent l'utilisation et la divulgation légales des informations de santé protégées aux États-Unis. Depuis 2018, en France, les hébergeurs de données de santé doivent passer la certification HDS pour garantir la qualité de service des hébergeurs de santé.

2) La blockchain

"*Une blockchain est un registre, une grande base de données qui a la particularité d'être partagée simultanément avec tous ses utilisateurs, tous également détenteurs de ce registre, et qui ont également tous la capacité d'y inscrire des données, selon des règles spécifiques fixées par un protocole informatique très bien sécurisé grâce à la cryptographie*." selon le rapport de la mission d'information commune sur la blockchain par **l'Assemblée nationale.**

La blockchain a été élaborée par une personne ou un groupe connu sous le nom de Satoshi Nakamoto. Elle est apparue en 2008, un an avant l'arrivée du bitcoin dont elle est la principale composante. Le but principal de la blockchain est de mettre en avant la décentralisation et la fiabilité des données.

Cette technologie fonctionne de la façon suivante, la blockchain est un registre, une grande base de données publiques qui stocke des transactions.

Il y a un accord public sur l'ordre des transactions sous forme de preuve de travail.

La preuve de travail est le système de sécurisation de la blockchain, également appelée système de minage. Les données sont déchiffrées et authentifiées par des mineurs.

Ensuite, les transactions sont ajoutées dans un bloc. Une fois le bloc validé, il est ajouté à la chaîne de bloc selon un temps défini, pour le Bitcoin, c'est toutes les 10 mn. Donc toutes les transactions des 10 dernières minutes sont ajoutées à la chaîne de bloc. D'où le nom blockchain.

La blockchain assure les fonctionnalités suivantes :

Immuabilité : car la blockchain ne peut être modifiée.

Authenticité : Les blocs ne peuvent être modifiés grâce au proof of work (preuve de travail).
Transparence : Tout le monde peut lire ce qu'il y a sur la blockchain, elle est accessible à tous.
Securite : Grace au proof of work
Non-répudiation (assure que l'émetteur du message ne pourra pas nier avoir émis le message dans le futur) : grâce à la présence des clés publique dans le hash

Depuis l'apparition des contrats intelligents (smart contrat) associés à la blockchain, cette technologie à un grand potentiel et multiple application possible. Il permet de sceller un engagement entre plusieurs parties par l'intermédiaire d'un code informatique sans l'intervention d'un tiers de confiance.

Toutes ces caractéristiques de la blockchain font de cette technologie un candidat pertinent pour répondre au besoin de confidentialité dans l'échange des dossiers de santé électronique.

3) Les DSE

Voici la définition des DSE / EHR selon le Bureau du coordonnateur national des technologies de l'information sur la santé américain [51].
Un dossier médical électronique (DME) est une version numérique du dossier papier d'un patient. Les DSE sont des dossiers en temps réel, centrés sur le patient, qui rendent l'information disponible instantanément et en toute sécurité aux utilisateurs autorisés. Bien qu'un DSE contienne les antécédents médicaux et thérapeutiques des patients, un système de DSE est conçu pour aller au-delà des données cliniques standard recueillies dans le bureau d'un fournisseur et peut inclure une vision plus large des soins d'un patient. Les DSE sont une partie essentielle des TI en matière de santé et peuvent :

- *Contenir les antécédents médicaux, les diagnostics, les médicaments, les plans de traitement, les dates d'immunisation, les allergies, les images radiologiques et les résultats de laboratoire et de test d'un patient*
- *Permettre l'accès à des outils fondés sur des données probantes que les prestataires peuvent utiliser pour prendre des décisions concernant les soins d'un patient*
- *Automatiser et rationaliser le flux de travail des prestataires*

III. Protocole de recherche

1) Définition de la revue littéraire systématique

Une revue systématique est une démarche qui utilise des méthodes pour chercher, sélectionner et synthétiser toutes les informations disponibles. Elle répond à une question de recherche clairement énoncée et indique explicitement les méthodes utilisées pour parvenir à la réponse. Ce qui fait la différence entre la revue systématique et d'autres types de revues, c'est que les méthodes de recherche sont conçues pour réduire les biais. Il y a plusieurs étapes à la rédaction d'une revue systématique :

1. Formuler une question de recherche
2. Développer un protocole
3. Rechercher toutes les études pertinentes
4. Appliquer les critères de sélection
5. Extraire les données
6. Synthétiser les données

Nous avons deux types de revue systématique, la revue littéraire systématique (SLR) et la revue d'étude de cartographie systématique (SMS). Voici une description de ces deux sous-catégories de la revue systématique.

Étude de cartographie systématique (SMS) : "Une large revue des études primaires dans un domaine de sujet spécifique qui vise à identifier les preuves disponibles sur le sujet [52]."

Revue systématique de la littérature (SLR) : "Une forme d'étude secondaire qui utilise une méthodologie bien définie pour identifier, analyser et interpréter toutes les preuves disponibles liées à une question de recherche spécifique d'une manière impartiale et (dans une certaine mesure) reproductible [52]."

Dans le cadre de notre problématique, le choix le plus pertinent se voit être celui de la méthode SLR étant donné la présence d'une question de recherche spécifique ainsi que le besoin de résultat pour chacun des papiers utilisés.

Pour réaliser une analyse documentaire de haute qualité, il est important de suivre une méthodologie solide. Ainsi, ce mémoire reprend les étapes de Kitchenham et al. [52] pour réaliser une revue systématique de la littérature (SLR). Cette tâche a été divisée en trois étapes principales comme suit :

1. **Planification :** au cours de cette phase, nous avons défini les sous questions et questions principales de recherche ainsi que les objectifs de notre SLR. En outre, les bases de données de la littérature qui serviront pour la recherche d'articles sont sélectionnés et des critères d'inclusion/exclusion sont définis.
2. **Réalisation** : la SLR est réalisée, en suivant le plan conçu précédemment. Les études sont extraites puis filtrées, et les articles restants sont lus. Un cadre

analytique est utilisé pour extraire les données nécessaires pour répondre aux questions de recherche.

3. **Rapport** : les résultats du SLR sont présentés de manière factuelle, ainsi qu'une évaluation de la qualité des études extraites. Ils sont ensuite discutés dans leur propre section.

2) Définition des questions de recherches

L'objectif de cette étude est d'extraire, à partir de la littérature, un ensemble de données sur l'importance accordée à la confidentialité dans les projets liés à la blockchain et aux dossiers de santé électronique (DSE). Pour atteindre cet objectif, voici la question de recherche centrale (QR) qui a été définie :

QR : Comment est abordée la notion de confidentialité dans l'échange de données médicales utilisant la blockchain ?

Pour examiner plus précisément la littérature, la question est scindée en trois sous questions de recherche. Elles nous permettront d'obtenir une analyse plus ciblée et de se rendre compte dans quelle mesure les exigences liées à la confidentialité / vie privée sont considérées.

La première sous-question cherche à identifier si la vie privée est définie de quelconques façons. Ceci afin d'avoir un premier ressenti sur la préoccupation des personnes pour ce terme impliquant plusieurs critères.

QR1 : Comment est définie la vie privée dans les papiers de recherches concernant l'échange de données médical utilisant la blockchain ?

La deuxième question s'intéresse au niveau de preuves fourni par les auteurs pour valider leur contribution pour la vie privée. C'est-à-dire les solutions mises en place pour répondre aux critères engendrés par ce concept de vie privée.

QR2 : Quel est le niveau de preuves fourni par les auteurs pour valider leur contribution pour la vie privée ?

Pour finir, la dernière question se concentre sur la provenance de ces solutions pour justifier leur bonne résolution aux critères de vie privée. Se rattache-t-elle à des notions légales, à une méthodologie ou doit-on se baser sur la parole des auteurs pour justifier leur participation au respect de la vie privée des patients pour leur solution ?

QR3 : Est-ce que ces pratiques se rattachent à des notions légales ?

En se basant sur nos connaissances et nos avis sur ces questions de recherche, nous avons émis des hypothèses de recherche présentées ci-dessous. Ce mémoire permettra de tester ces hypothèses et de valider ou invalider celles-ci.

HR1 : La notion de vie privée n'est pas clairement, si ce n'est pas défini. Il faudrait se limiter à croire les chercheurs sur parole quand ils mentionnent le respect de la confidentialité dans leurs études.

HR2 : Le niveau de preuve de respect de la vie privée apporté par les auteurs n'est pas suffisant.

HR3 : La notion de vie privée dans les papiers de recherche n'est pas axée sur des concepts légaux. Ces propriétés ne paraissent pas vérifiées ni même conformes à des concepts légaux.

Maintenant que nous avons déterminé les questions de recherche qui vont nous guider dans l'analyse de cette revue et permettre d'identifier les études primaires pertinentes. Nous allons par la suite définir les stratégies de recherches.

3) La stratégie de la recherche

La stratégie de recherche de notre revue systématique de littérature a pour objectif de trouver de nombreuses publications scientifiques s'exprimant sur notre problématique et ensuite appliquer les critères de sélection exclusion afin de ne garder uniquement les publications pertinentes dans le cadre de notre SLR.

La blockchain étant une technologie récente et complexe, nous avons fait le choix de nous concentrer sur les papiers dont les architectures sont clairement définies, et avons mis par conséquent un critère de sélection sur l'utilisation de pattern architectural, qui sont un gage de qualité de la description architecturale.

De plus, parallèlement à cette étude, une SLR est conçue sur ce sujet, l'identification des patterns logiciels architecturaux basés sur la blockchain qui sont utilisés dans les systèmes actuels de dossier de santé électronique DSE.

Afin de correctement documenter l'ensemble du processus et de pouvoir travailler conjointement sur la méthodologie de recherche et ses résultats, nous avons utilisé Parsifal, un outil en ligne conçu pour aider les chercheurs à effectuer des revues de littérature systématiques dans le contexte du génie logiciel . Les chercheurs répartis géographiquement peuvent travailler ensemble dans un espace de travail partagé, concevoir le protocole et mener la recherche [53].

De plus, cet outil est une aide pour se rappeler ce qui est important lors d'une revue systématique de la littérature. Pendant la phase de planification, Parsifal fait le rappel avec

les objectifs, le PICOC, les questions de recherche, la chaîne de recherche, les mots-clés et les synonymes, la sélection des sources, les critères d'inclusion et d'exclusion et fournit également des mécanismes pour créer une liste de contrôle d'évaluation de la qualité et des formulaires d'extraction de données [53].

Ainsi, nous avons rapproché nos méthodologies de recherches avec l'objectif de nous concentrer sur une même base d'articles puis de traiter le résultat sous un angle différent pour répondre à nos problématiques succinctes.

Ensuite, sur la base de ces recherches pertinentes, nous avons effectué la recherche en boule de neige, aussi nommée "Snowballing" souvent utilisée dans le domaine de la recherche en sociologie et en statistique, correspondant à l'échantillonnage boule de neige pour étendre notre base d'articles.

a. Définition de la requête de recherche

La première consiste à établir les mots clés qui nous serviront à concevoir notre requête de recherche. Ces termes associés de la bonne façon nous permettront d'obtenir notre base de publications.

Les mots clés ayant été retenus pour construire la requête sont les suivants :
"Privacy", "Blockchain", "Architecture", "Patterns", "Healthcare", Electronic records".
Afin d'élargir le résultat de la requête, nous avons également introduit leurs synonymes correspondants.

Enfin, les connecteurs logiques, opérateurs booléens, AND et OR ont été utilisés pour relier les mots clés et y associer leurs synonymes.

Voici donc la requête obtenue à la finalité :

("Electronic Records" OR "Electronic Files") AND ("blockchain") AND ("healthcare" OR "health" OR "health-care" OR "medical") AND ("patterns" OR "architecture" OR "design" OR "model") OR ("privacy" OR "confidentiality" OR "private" OR "protection" OR "security")

La source de publication utilisée a été MIAGE Scholar qui a pour propriété d'exporter les données de la littérature dans le cadre de la réalisation de SMS / SLR. La recherche en boule de neige à partir des articles sélectionnés est également considérée comme une source de données, car elle peut permettre d'inclure d'autres articles pertinents.

Sur MIAGE Scholar la requête précédente équivaut à celle ci-dessous :

(TITLE(healthcare) OR TITLE(health) OR TITLE(health-care) OR TITLE(medical)) AND (TITLE(blockchain)) AND (TITLE(electronic files) OR TITLE(electronic records)) AND (TITLE-ABS-KEY("patterns") OR TITLE-ABS-KEY("architecture") OR

TITLE-ABS-KEY("design") OR TITLE-ABS-KEY("model")) AND (TITLE-ABS-KEY("privacy") OR TITLE-ABS-KEY("confidentiality") OR TITLE-ABS-KEY("private") OR TITLE-ABS-KEY("protection") OR TITLE-ABS-KEY("security")) AND PUBYEAR < 2022

Un critère “PUBYEAR < 2022” sur la date de publication à été choisi pour toujours obtenir le même nombre de résultats à l'exécution de la requête.

b. Liste de contrôle pour l'évaluation de la qualité

Durant la deuxième phase de planification de la recherche, pour évaluer la qualité méthodologique des revues systématiques, nous avons mis en place une liste de questions d’évaluation (Quality Assessment QQ) ci-contre :

QQ1 : La notion de privacy est-elle clairement définie ?

Afin de tester les hypothèses énoncées précédemment, s’agissant de savoir si la gestion de la confidentialité / vie privée est bien prise en considération dans les publications existantes.

QQ2 : La notion de privacy est-elle rattachée à des notions légales ?

Il s’agit ici de déterminer si des textes de loi sont mentionnés, utilisés, pour appuyer les dires des auteurs.

QQ3 : Est-ce que la solution est en conformité avec les recommandations légales ?

Nous cherchons à savoir si la solution propose concrètement des dispositions en conformité avec les recommandations légales.

QQ4 : Est-ce que le papier fait partie d'une bonne conférence ?

Un papier faisant partie d'une bonne conférence est essentiel pour confirmer ou pas notre démarche et nos hypothèses. Cela est très important également dans le cadre de l'authenticité de cette SLR.

Chaque question a respectivement trois propositions de réponse :

- “oui”avec un poids de 1
- “partiellement” avec un poids de 0.5
- “non” avec un poids de 0

En rejoignant les questions de la seconde SLR portant sur l'identification des patterns logiciels architecturaux, nous obtenons sept questions. Par conséquent, le score peut être compris entre 0 et 7.

La réponse à ces questions nous permet d'obtenir le score d'évaluation de la qualité. Cela permet d'évaluer globalement la qualité de l'article par rapport à nos objectifs.

Utilisation des QQ dans le processus de sélection des articles

Les articles obtiennent en majorité une note inférieure à la moyenne qui est de 3,5/7.

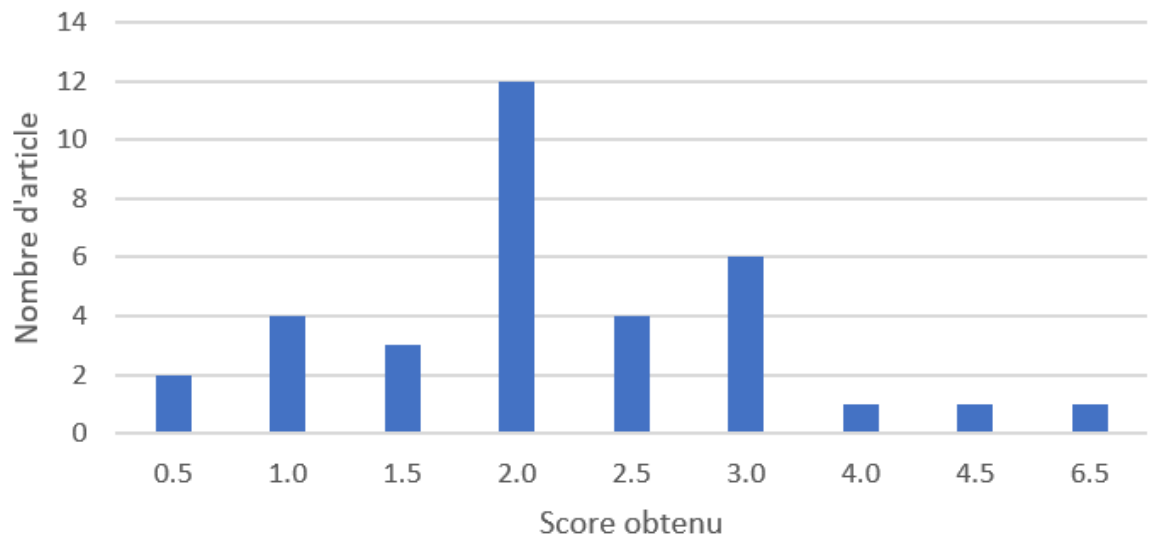


Figure 1 : Somme des scores d'évaluation de la qualité obtenue

La moyenne des scores est de 2,25/7 et seulement trois articles ont obtenu un résultat supérieur à la moyenne.

Cependant, nous avons choisi de ne pas utiliser le score d'évaluation de la qualité pour sélectionner un article dans le cadre de nos SLR. Nous n'avons pas utilisé les Quality Assessment QQ comme il est d'usage, car l'objectif de notre SLR est justement de juger de la qualité des papiers.

En effet, nos articles étant déjà sélectionné via nos critères d'inclusion et d'exclusion que nous verrons ci-dessous, notre approche pour ces questions cherche à indiquer, dans le cadre de ma SLR, si la notion de privacy est abordée au sein des publications et quel est le niveau de preuve fourni par celles-ci.

Toutes les références au terme "privacy" se basent sur la définition donnée en section II "Définition des termes clés".

c. Critères d'inclusion / exclusion

Pour la stratégie de sélection, il faut effectuer la phase de filtrage de la SLR. Des critères d'inclusion et d'exclusion ont été mis en place. Ils permettent de fournir des directives systématiques pour inclure ou exclure des articles pendant la phase de filtrage, où les articles sont sélectionnés pour une lecture plus approfondie.

Le tableau 1 présente les critères d'inclusion et d'exclusion retenus :

Tableau 1 : Critères d'inclusion et d'exclusion

Critères d'inclusion	Critères d'exclusion
1. Contient une architecture de dossier de santé électronique utilisant la blockchain	1. Architecture utilisant la blockchain non définie 2. L'article est une copie d'autres études. 3. L'article est une revue de littérature 4. L'article n'est pas en anglais. 5. L'article n'est pas accessible 6. L'article est trop ancien.

On peut remarquer qu'il n'y pas de critères d'inclusion et d'exclusion mentionnant la vie privée ou la confidentialité. Cela se justifie en relisant notre hypothèse **HR1** qui indique que la vie privée n'est pas réellement définie par les chercheurs dans leurs articles. Il ne serait donc pas propice de mettre un critère d'inclusion ou d'exclusion sur cette spécificité.

Critères d'inclusion :

La majorité des articles contenant une architecture de dossier de santé électronique utilisant la blockchain ont été sélectionnés dans le cadre de notre besoin d'analyser les contributions dont les architectures sont clairement définies avec une utilisation de pattern architectural, gage de qualité de la description architecturale.

Critères d'exclusion :

Les publications ne présentant pas une architecture utilisant la blockchain définie ni même détectable n'ont pas le besoin d'être traitées dans le cadre de notre problématique.

Les revues systématiques sont une synthèse des connaissances existantes sur une problématique, ce n'est pas ce que nous cherchons à examiner ici.

Nous avons choisi de ne traiter que les publications en anglais. De plus, la requête a été formulée en anglais afin d'obtenir un meilleur résultat.

Exécution de la requête

Pour rappel, voici ci-dessous la requête utilisée pour obtenir la liste des articles :

(TITLE(healthcare) OR TITLE(health) OR TITLE(health-care) OR TITLE(medical)) AND (TITLE(blockchain)) AND (TITLE(electronic files) OR TITLE(electronic records)) AND (TITLE-ABS-KEY("patterns") OR TITLE-ABS-KEY("architecture") OR TITLE-ABS-KEY("design") OR TITLE-ABS-KEY("model")) AND (TITLE-ABS-KEY("privacy") OR TITLE-ABS-KEY("confidentiality") OR TITLE-ABS-KEY("private") OR TITLE-ABS-KEY("protection") OR TITLE-ABS-KEY("security")) AND PUBYEAR < 2022

Cette requête nous a permis de récupérer 59 publications. Nous avons ensuite appliqué les critères d'inclusion exclusion sur ces publications. Cela nous a conduits à retirer 23 articles.

Voici le déroulé du processus :

- Nous avons recensé sept publications n'étant pas rédigées en anglais ou inaccessibles (critère quatre et cinq). Le critère deux concernant les copies d'autres études a été concerné par une unique publication.
- Nous avons poursuivi par la lecture des publications en lisant le titre de l'article, l'abstract, l'architecture mise en place ainsi que la présence de référence à la vie privée. Cette première analyse nous a fait enlever encore cinq publications de notre base.
- La suite de l'analyse consistant en la lecture des quarante-six publications restantes s'est conclue par l'exclusion de dix publications.
- L'ajout des publications par la méthode de snowballing nous a permis de retenir deux publications supplémentaires. Le reste n'ayant pas été retenue car :
    - Privacy non mentionnée
    - L'article ne faisant pas référence :
        - Au dossier de santé électronique (DSE)
        - À la blockchain
    - Pas d'information supplémentaire que celles fournies par la publication "parente".

Cela nous fait donc un total final de trente-huit publications, la figure ci-dessous illustre les phases d'analyses et de filtrages appliquées sur notre base de publication pour arriver à une finalité de trente-six publications.

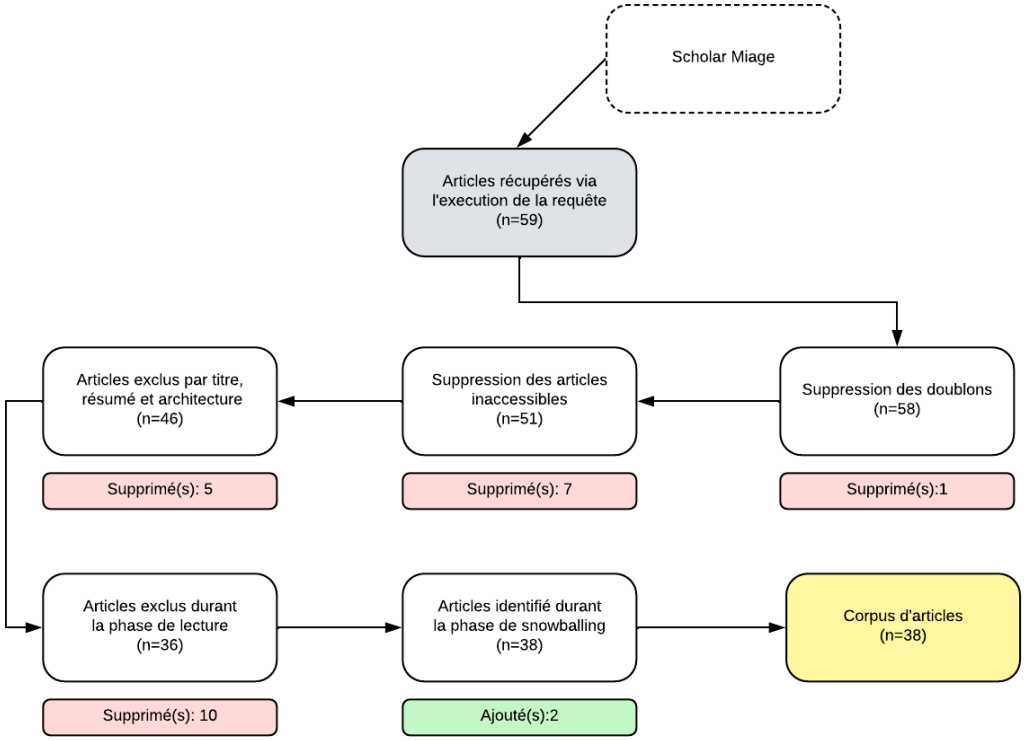


Figure 2 : Schéma du processus de révision.

Les articles ajoutés via la méthode “snowballing” étant des revues n'auraient pas dû être ajoutés. Ils ont été sélectionnés uniquement, car ils ont certains éléments liés à la privacy qui représentent bien ce que l’on devrait retrouver dans une publication de la littérature scientifique.

IV. Extraction et analyse des données

La section quatre de notre revue systématique de littérature présente une analyse quantitative et qualitative sur les données obtenues.

La première partie, l'analyse quantitative, est une analyse statistique des résultats obtenus sur l'extraction des données concernant l'ensemble des publications. Cette analyse nous permettra d'avoir une vision d'ensemble de la littérature scientifique sur notre problématique.

La seconde partie, l'analyse qualitative, permettra de répondre de manière plus précise à nos interrogations. Cela s'effectuera en se concentrant sur un article en particulier avec l'utilisation d'un outil permettant de modéliser des menaces à la vie privée. Cet outil fournit un soutien méthodologique et des connaissances étendues pour aborder systématiquement l'analyse des menaces à la vie privée.

1) Analyse quantitative

Comme pour l'application de la stratégie de recherche, nous avons utilisé parsif.al pour l'extraction et l'analyse des données. Nous avons répondu aux différentes questions mises en place pour l'évaluation de la qualité des publications directement sur parsif.al pour l'ensemble des 36 articles. Cela nous a permis de générer un tableau nommé "data extraction form" listant la réponse à ces questions pour chaque publication.

C'est ainsi que nous obtenons les différentes analyses statistiques sur notre base de publications. Cela nous a permis d'avoir en grande partie une réponse à nos hypothèses émises au début de notre revue systématique de littérature. Également, cela nous permet d'identifier les données utiles que nous pouvons extraire des publications.

Pour rappel, voici les questions pour l'évaluation de la qualité ainsi que leurs options de réponses possibles :

QQ1 : La notion de privacy est-elle clairement définie ?
La réponse attendue est un booléen, VRAI ou FAUX

QQ2 : La notion de privacy est-elle rattachée à des notions légales ?
La réponse attendue est un booléen, VRAI ou FAUX

QQ3 : Est-ce que la solution est en conformité avec les recommandations légales ?
La réponse attendue est un booléen, VRAI ou FAUX

QQ4 : Si oui. La solution permet de se conformer aux recommandations légales de privacy suivante :

- Les données patientes collectées doivent être conservées pour une durée déterminée

- L’accès aux données de santé des patients est / peut être limité
- Prioriser l’utilisation de blockchain à permission
- Respecter des règles de sécurité pour protéger les données des patients contre des accès non autorisés (ex : utilisation de la carte professionnelle de santé, mot de passe personnel, utilisation d’un système de chiffrement fort, etc)
- Tenir un registre des activités fiables de traitement et le renseigner

QQ5 : Est-ce que le papier fait partie d'une bonne conférence ?
- Conf A journal Q1 (very best venues)
- Conf B journal Q2 (good venues)
- rest (least ranked venues or unranked)

A) La notion de privacy est-elle clairement définie ?

Comme nous pouvons le voir sur la figure ci-dessous, la quasi-totalité des publications ne définissent pas explicitement la privacy :

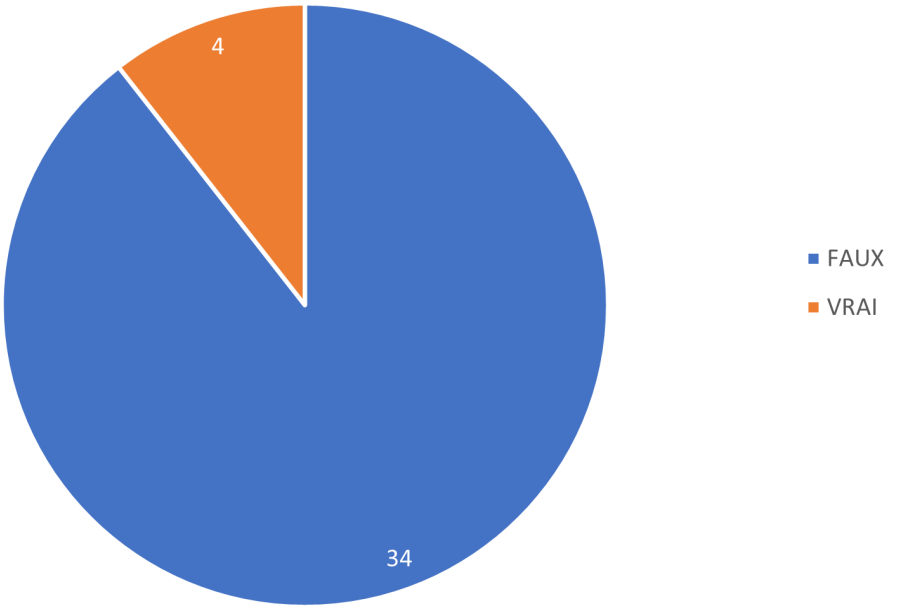


Figure 3 : La notion de privacy est-elle clairement définie ?

Tout d’abord, clarifions les choses en regardant ce que signifie “clairement définie". Cela implique que le terme de privacy est explicitement défini au sein de la publication et qu’une définition telle que nous l’avons faite en section II est donnée par l’auteur. Il se peut qu’elle ne soit pas définie dans toute sa complexité, mais cela prouve au moins que la publication s’est intéressée à cette notion.

Sur un total de 38 publications, 34 publications ne **définissent pas clairement la privacy**. On peut visualiser ce résultat en deux sous-catégories. Les publications qui ne mentionnent trop peu voir nullement la privacy et celles qui la mentionnent sans vraiment en donner une définition.

Nous pouvons, par exemple, identifier certains articles qui ne mentionnent aucunement la privacy pour leurs solutions [21] [8] [34]. Parmi ce genre d'article, le terme privacy est utilisé en introduction pour dire que la blockchain est une solution pour assurer la confidentialité, la sécurité et la vie privée (privacy) [21]. Il y a également des articles qui ne rapportent pas du tout ni même le simple mot privacy dans le contenu de leur solution [34].

Le type de publications déclaré ci-dessus constitue une partie minoritaire de notre base d'étude. La majorité des publications utilisent le terme de vie privée / privacy mais ne le définissent pas et donc ne déclare la complexité engendrée quand on veut qu'une solution respecte la notion de privacy. La plupart des articles affirment que leurs solutions respectent la privacy sans donner de justification.

Il y a les articles qui ne justifient clairement pas le respect de la privacy grâce à leur solution. Voici pour exemple la conclusion d'un article de ce type :
[12] : *"Dans cet article, nous avons proposé un mécanisme basé sur la blockchain pour protéger la confidentialité des données. Dans le système proposé, les protocoles et algorithmes de communication et d'authentification entre les entités n'ont pas été entièrement étudiés"*

Un seul article de cette catégorie ne définit pas clairement la privacy mais semble tenir compte de ses exigences, car il fait mention de la réglementation générale sur la protection des données RGPD et l'HIPAA et respecte certaines règles liées à celle-ci de façon explicite [27].

En général, nous constatons que les publications respectent certaines règles liées à la privacy de manière non explicite, sans justification par une quelconque réglementation ou lois existantes. Nous verrons cela plus en détail en analysant les résultats des questions **QQ4** et **QQ5**.

Les rares publications qui **définissent la privacy sont au nombre de 4**, la moitié issue du « snowballing » étant incluses pour des raisons d'illustration de ce qu'il serait mieux de faire concernant la privacy.

Les deux autres articles qui définissent la privacy se partagent en deux catégories. Un article défini correctement la privacy et le second la définie selon les propos de l'auteur qui ne correspondent pas à ce que l'on devrait s'attendre quand on parle de privacy :
[27] : *"La confidentialité : Quelles données peuvent être partagées avec qui, et comment les conditions médicales jugées sensibles par le patient (par exemple, les addictions) peuvent être traitées."*

[30] : *“Confidentialité des données : un réseau blockchain est un grand livre distribué dans lequel tous les acteurs du secteur de la santé, y compris les patients, stockent leurs principales données médicales électroniques, et tout le monde a accès à ces données privées sensibles sur la plateforme. Cela crée de sérieux problèmes de confidentialité car la majorité des patients et des autres parties prenantes ne veulent pas utiliser leurs données privées contre leurs concurrents…”*
Nous pouvons donc dès maintenant constater que notre première hypothèse **HR1** semble se confirmer.

De ce que l’on a obtenu depuis notre base de publication, seuls les deux articles récupérés via la méthode de snowballing nous ont permis de récupérer un contenu plus intéressant.

Ils décrivent précisément la privacy comme nous pouvons le voir ci-dessous :

[37] : *“La vie privée, ou la confidentialité des informations, pour être plus précis, fait référence aux informations personnelles d'un patient qui sont collectées pour aider à identifier un individu. La sécurité, quant à elle, consiste à restreindre et à autoriser l'accès aux informations personnelles. Ces deux aspects sont essentiels, car la fuite d'informations peut avoir de graves conséquences. “*

[38] : *« 1) VIE PRIVÉE*
La vie privée désigne le droit qu'a une personne de décider quand, comment et à quels niveaux elle peut accéder à ses DSE personnels, les transformer et les partager avec d’autres... La vie privée peut être violée dans diverses situations ; par exemple, un fournisseur de soins de santé peut abuser des DSE, intentionnellement ou par erreur [41]. Dans un document d'enquête, Win [42] indique qu'environ deux tiers des patients prêtent attention à leurs DSE personnels. Dans une autre enquête, Ancker et al. [43] ont indiqué que près de cinquante pour cent des participants pensent que l'échange de données de santé aggraverait la confidentialité de leurs données. La confidentialité est donc un facteur important à prendre en compte lors de la comparaison des solutions basées sur la blockchain qui prétendent préserver la confidentialité des DSE. »

B) La notion de privacy est-elle rattachée à des notions légales ? Si oui, lesquelles ?

Intéressons-nous désormais au fait de savoir si les publications font mention de lois, réglementation ou standard portant sur la protection de la vie privée tel que les éléments suivants :

- Health Insurance Portability and Accountability Act (HIPAA) : Loi fédérale américaine qui exige la création de normes nationales pour protéger les informations sensibles sur la santé. De cette loi découlent les règles de confidentialité HIPAA pour mettre en œuvre les exigences de l'HIPAA.
- Règlement général sur la protection des données (RGPD) : Réglementation encadrant le traitement des données personnelles sur le territoire de l'Union européenne
- Fast Healthcare Interoperability Resources (FHIR) : standard d'interopérabilité pour les échanges des informations médicales

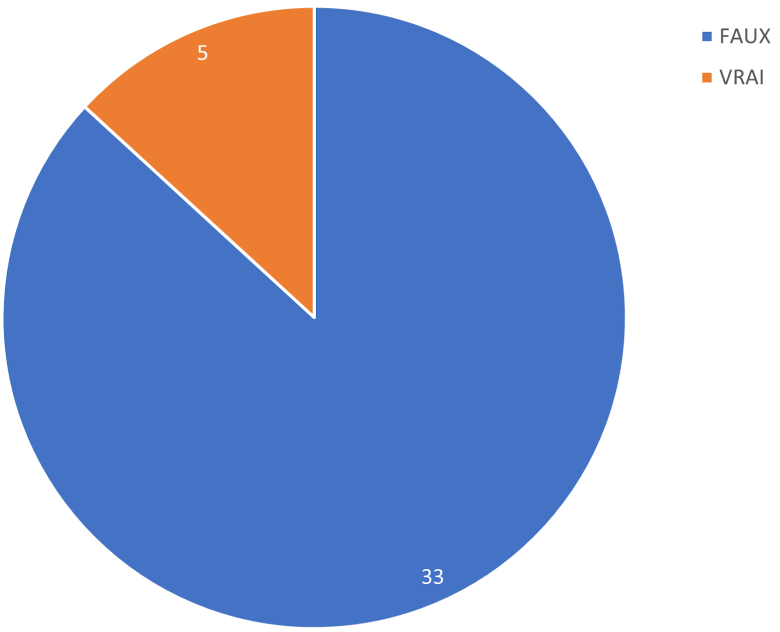


Figure 4 : La notion de privacy est-elle reliée à des notions légales ?

Ce graphique nous permet de constater que les publications ne mentionnent en majorité aucune dispositions existantes régissant sur la privacy des données, tout comme elles ne la définissent pas clairement. Cela peut donc laisser paraître que les auteurs ne connaissent pas ou ne prennent pas en considération les recommandations et exigences légales dans leurs solutions.

Comme nous pouvons le constater, nous avons plus d'articles faisant référence à des notions légales que des articles qui définissent la privacy, cinq articles contre quatre qui définissent la privacy.

Nous avons en effet constaté plusieurs types d'articles selon ces caractéristiques :

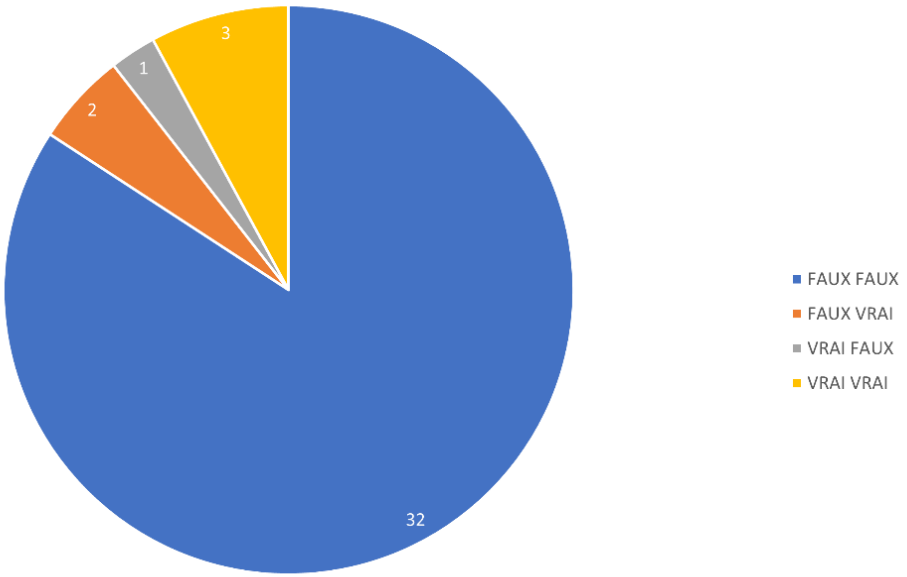


Figure 5 : Répartition des articles dont la privacy est définie et reliée à des notions légales

Nous pouvons distinguer plusieurs situations :

1. La publication définit la privacy et là relie des notions légales
2. La publication définit la privacy et ne la relie à aucune notion légale
3. La publication ne définit pas la privacy et la relie à des notions légales
4. La publication ne définit pas la privacy et ne la relie à aucune notion légale

Le premier cas de figure représente 3 articles au sein de notre base, soit 6% de la totalité des articles, tout en sachant que nous avons deux publications issues du snowballing étant des revues de la littérature. Nous avons donc une seule publication dans notre base qui définit la privacy tout en mentionnant un texte de loi [30].

[21] : *« L'architecture Hyperledger Fabric proposée est extrêmement conforme à la loi HIPAA (Health Insurance Portability and Accountability Act) et à la norme ISO/TS 18308 [14]. »*

Le deuxième concernant le type de publication qui définit la privacy sans la relier à des notions légales est concerné par un article dans notre base [15]. L'article en question ce souci cependant de la juridiction au niveau des soins de santé, mais cet intérêt ne semble pas se diriger vers la privacy des données :

[15] : « Dans le cadre de ce travail, l'espace problématique sera limité aux exigences des juridictions de soins de santé réglementées responsables devant les autorités gouvernementales (c'est-à-dire les organismes de santé publique et d'assurance sociale), qui, au niveau le plus fondamental, maintiennent un registre à l'échelle de la juridiction bénéficiaires et prestataires dans le but de rémunérer les prestataires pour les services de santé offerts aux bénéficiaires. Nous examinons le scénario de plus en plus courant où ces autorités agissent principalement en tant que régulateurs des services de soins de santé plutôt que de fournir des services de santé »

Le troisième concerne deux publications de notre base. Il introduit le fait que même si certains articles ne définissent pas la privacy, ce n'est pas pour autant par le fait qu'ils ne la prennent pas en considération dans leur solution.

[27] : *« HRC introduit un cadre amélioré pour garantir la sécurité et la confidentialité des utilisateurs et se conformer aux réglementations connexes (Règlement général sur la protection des données GDPR, Health Insurance Portability and Accountability Act HIPPA) en utilisant la blockchain. »*

[26] : *« De plus, dans la conception de l'architecture technologique, deux réglementations internationales et une loi péruvienne sont prises en compte pour s'assurer que l'architecture technologique répond aux exigences liées à la gestion des données médicales :*
1) La règle de confidentialité HIPAA : Protège la vie privée des patients et vous donne un meilleur accès à votre dossier médical en exigeant de toutes les personnes concernées qu'elles protègent la confidentialité des informations de santé des patients en veillant à ce que les dossiers médicaux ne soient disponibles que lorsque cela est nécessaire [27].
2) Règlement général sur la protection des données (RGPD) :
Intensifie et unifie la protection des données à l'intérieur et à l'extérieur de l'UE afin d'offrir aux utilisateurs le contrôle de leurs données personnelles et de simplifier l'environnement réglementaire [28].
3) Loi sur la protection des données personnelles (N° 29733) :
Loi péruvienne qui a été créée pour protéger les données personnelles des personnes physiques, des entités juridiques ou des entités publiques, ce qui englobe toutes les informations qui rendent une personne identifiable,
et fixer des normes pour la protection des informations sensibles en établissant des mécanismes pour leur contrôle [29]. »

De plus, on constate que toutes les publications de notre base, l'ayant définis ou pas, qui ont relié la privacy avec des notions légales, ont exprimé explicitement le fait que leur solution est en conformité avec ces réglementations mentionnées.

Le dernier cas de figure est de loin le cas majoritaire. On dénombre en effet 32 articles dans ce cas, soit 67% de notre base de publications. Cela s'inscrit dans la continuité de notre hypothèse qui tend à dire que les auteurs ne se soucient pas assez de la privacy dans leurs études.
Il y a certaines études qui font mention de certaines réglementations, mais nous ne pouvons pas considérer cela comme si l'article reliait la privacy à ces réglementations.

Par exemple, l'article ci-dessous mentionne l'HIPAA une seule fois dans son introduction pour donner une information :

[29] : "*Selon la loi sur la portabilité et la responsabilité en matière d'assurance maladie (HIPAA) en Europe, les incidents liés aux soins de santé les incidents de piratage ont augmenté de 42 % en 2020, poursuivant une tendance de 5 ans qui a vu les incidents de piratage augmenter chaque année.* "

La réglementation la plus citée par les publications est la Health Insurance Portability and Accountability Act (HIPAA) suivie par la réglementation de l'Union Européenne, le Règlement général sur la protection des données (RGPD) :

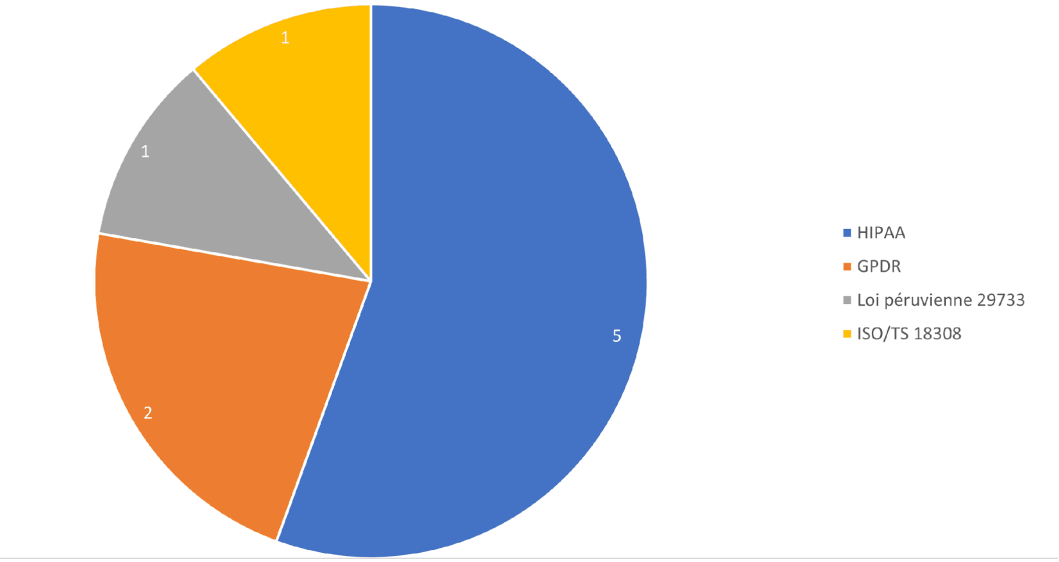


Figure 6 : Quels sont les réglementations citées dans les publications ?

Ce résultat peut s'expliquer par le fait que l'HIPAA régit directement sur les données de santé.

Ces informations approuvent nos hypothèses HR2 et HR3 qui indiquent que le niveau de preuve fourni par les auteurs n'est pas suffisant ni axé sur des notions légales. Maintenant que nous avons établi qu'en majorité les auteurs proposant des solutions pour l'échange de données de santé utilisant la technologie blockchain ne s'intéressent pas suffisamment, voire aucunement, à la privacy et aux réglementations sous-jacentes, nous allons nous intéresser au fait de savoir si ces solutions sont tout de même en adéquation avec ces réglementations et leurs recommandations.

C) Les solutions sont-elles en conformités avec les recommandations légales ?

On constate grâce à nos analyses que la privacy dans toute sa complexité n'est pas vraiment prise en considération dans le traitement des données de santés utilisant la blockchain. Cela est problématique et peut constituer un frein au développement d'une solution blockchain, sachant que les données concernant la santé sont d'autant plus contrôlées, car extrêmement sensibles. Comme nous l'avons vu, des réglementations ont été mises en place dans l'unique but d'encadrer la gestion des données sensibles comme les données de santé.

Nous avons constaté que les publications reliant la privacy à ces réglementations expriment explicitement le fait que leur solution est compatible avec celles-ci. Mais nous pouvons nous poser la question suivante, est-ce que le fait que les auteurs des publications ne s'impliquent pas suffisamment sur la privacy, ne la décrivent pas, ne s'inquiètent pas assez de cette problématique, veut-il forcément sous-entendre que les solutions mises en place ne la respectent pas ?

Nous allons voir dans cette partie que la non-considération de la privacy au sein d'une publication ne veut pas systématiquement indiquer que la solution proposée ne la respecte pas. En effet, les deux ne sont pas forcément imbriqués.

Comme nous le voyons ci-dessous, 87% des articles de notre base sont en conformité avec des recommandations légales.

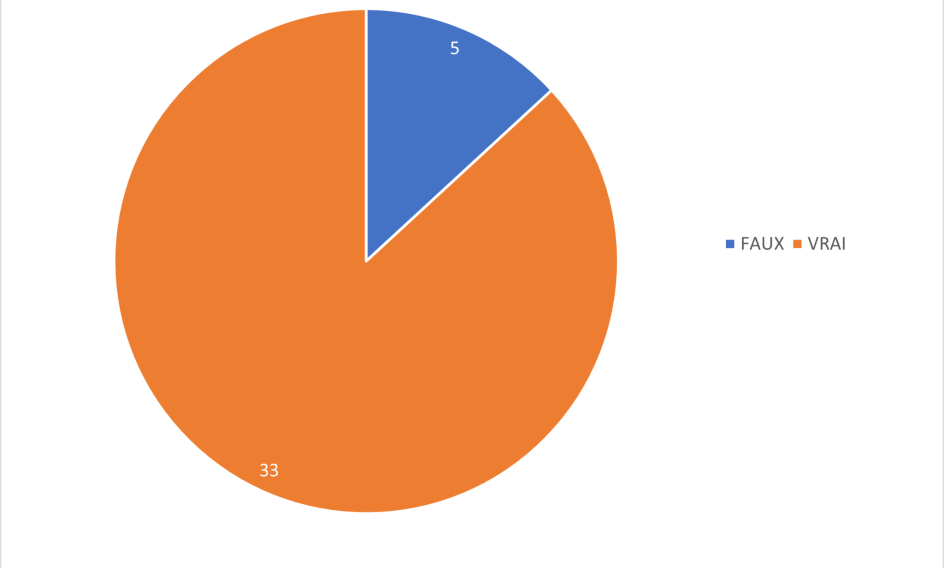


Figure 7 : Est-ce que la solution est en conformité avec des recommandations légale ?

Le résultat à cette question est déterminé en analysant le respect de la solution avec les mesures mises en place avec le RGPD. Pour rappel, il y a cinq principales mesures pour régir l'encadrement des données en termes de santé :

- L'accès aux données de santé des patients est / peut être limité
- Prioriser l'utilisation de blockchain à permission
- Respecter des règles de sécurité pour protéger les données des patients contre des accès non autorisés (ex : utilisation de la carte professionnelle de santé, mot de passe personnel, utilisation d'un système de chiffrement fort, etc)
- Tenir un registre des activités fiables de traitement et le renseigner
- Les données patients collectées doivent être conservées pour une durée déterminée (droit à l'oubli)

Le fait que les solutions proposées respectent en majorité ces mesures s'explique par la technologie qu'elles utilisent. En effet, la technologie blockchain possède un grand avantage dans le respect d'une partie des mesures exprimées ci-dessus. Par défaut, grâce à ses caractéristiques, la blockchain est une bonne candidate en termes de privacy.

La CNIL exprime les bénéfices de cette technologie dans le cas des données de santé [54] :

- L'immuabilité des actions effectuées sur la Blockchain permet le développement de solutions permettant de répondre aux obligations de traçabilité du consentement ou des actions effectuées sur les données.
- S'agissant de l'exercice des droits, certains droits peuvent être exercés de manière effective tels que le droit d'accès et le droit à la portabilité.
- Sans pouvoir conduire à des effets strictement identiques, ces solutions permettent de se rapprocher des exigences de conformité du RGPD, notamment en coupant l'accessibilité de la donnée en fonction du format choisi

Par ailleurs, les principes en matière de sécurité demeurent entièrement applicables dans la Blockchain.

Les caractéristiques de la blockchain apportent aux solutions une prédisposition positive aux questions sur la privacy des données. Cela explique le fait que 87% des publications soient en conformité avec des recommandations exprimées par le RGPD.

Certaines recommandations sont plus faciles à appliquer que d'autres. Nous allons voir la répartition des recommandations les plus respectées dans les articles. Ces réponses sont basées sur ce que nous avons pu constater.

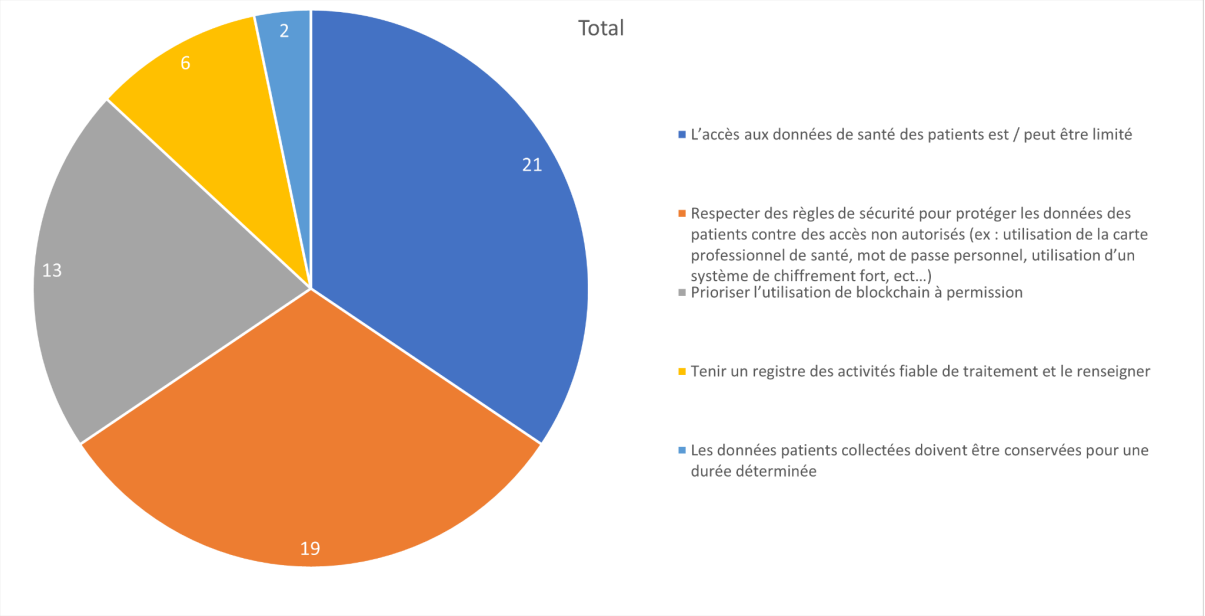


Figure 8 : La solution permet de se conformer aux recommandations légales de privacy suivantes

Trois recommandations sont largement respectées de façon explicitent dans les publications. Il s'agit des recommandations suivantes :

- L'accès aux données de santé des patients est / peut être limité
- Prioriser l'utilisation de blockchain à permission
- Respecter des règles de sécurité pour protéger les données des patients contre des accès non autorisés (ex : utilisation de la carte professionnelle de santé, mot de passe personnel, utilisation d'un système de chiffrement fort, etc)

Nous rentrerons ultérieurement plus en détail sur la blockchain à permission mais on remarquera que la recommandation indiquant d'utiliser une blockchain à permission et celle indiquant que l'accès aux données de santé des patients doit être limité sont concordantes entre elles. Il apparaît donc logique que si l'une est respectée en majorité, l'autre le soit également.

La recommandation concernant le respect des règles de sécurité pour protéger des accès non autorisés est respectée en grande majorité de manière explicite. C'est souvent via cet unique dispositif que les auteurs justifient la sécurité et la privacy des données.

Au sujet de la recommandation pour le registre des activités, elle est très peu évoquée dans les articles. Il est en majorité évoqué par les publications qui font références à des réglementations en vigueur.
Rappelons toutefois que par définition, la blockchain est un registre, une grande base de données qui a la particularité d'être partagée simultanément avec tous ses utilisateurs.

On peut donc estimer que malgré qu'il soit peu évoqué dans les articles, le respect de cette recommandation est induit par la simple utilisation de la blockchain. Même si sans disposition préalable, le traçage des activités peut s'avérer un peu complexe.

Le point le plus problématique avec l'utilisation de la blockchain concerne la recommandation précisant que les données de santé doivent être conservées pour une durée déterminée pour respecter le droit à l'effacement, également appelé droit à l'oubli.

Art. 17 de la RGPD [55] :

"1. La personne concernée a le droit d'obtenir du responsable du traitement l'effacement, dans les meilleurs délais, de données à caractère personnel la concernant et le responsable du traitement a l'obligation d'effacer ces données à caractère personnel dans les meilleurs délais, lorsque l'un des motifs suivants s'applique :

a) les données à caractère personnel ne sont plus nécessaires au regard des finalités pour lesquelles elles ont été collectées ou traitées d'une autre manière ;

b) la personne concernée retire le consentement sur lequel est fondé le traitement, conformément à l'article 6, paragraphe 1, point a), ou à l'article 9, paragraphe 2, point a), et il n'existe pas d'autre fondement juridique au traitement ;

c) la personne concernée s'oppose au traitement en vertu de l'article 21, paragraphe 1, et il n'existe pas de motif légitime impérieux pour le traitement, ou la personne concernée s'oppose au traitement en vertu de l'article 21, paragraphe 2 ;

d) les données à caractère personnel ont fait l'objet d'un traitement illicite ;

e) les données à caractère personnel doivent être effacées pour respecter une obligation légale qui est prévue par le droit de l'Union ou par le droit de l'État membre auquel le responsable du traitement est soumis ;

f) les données à caractère personnel ont été collectées dans le cadre de l'offre de services de la société de l'information visée à l'article 8, paragraphe 1."

En effet, la blockchain ayant les caractéristiques d'être authentique et immuable, les données stockées au sein des transactions dans une blockchain ne peuvent être effacées.

Cela est en parfaite opposition avec la règle du droit à l'effacement évoqué par le RGPD. Cela explique en conséquence le fait que la recommandation concernant ce droit soit si peu respectée par les solutions proposées.

Deux publications proposent des solutions pour pallier cette problématique. La première fonctionne par l'intermédiaire et l'utilisation du pattern **off-chain data storage** :

“Le RGPD donne au patient le droit de supprimer tous ses dossiers médicaux. Ainsi, lorsque nous sauvegardons des données hors chaîne, nous pouvons supprimer la valeur de hachage, et l'adresse du portefeuille sera supprimée. Le patient contrôle ses données de DSE, il sait où elles seront enregistrées et qui y aura accès, et il doit autoriser la mise à jour des données.” [19]

Le pattern **off-chain data storage** est principalement utilisé pour stocker de grandes quantités de données au sein de la blockchain. Il est utile, car la blockchain a une capacité de stockage limitée. Quand on veut profiter de l’immuabilité et l’intégrité sur des données d’un volume important, ce pattern est souvent utilisé. Il consiste à stocker une valeur de hachage des données brutes sur la blockchain. Le hachage permet de d’obtenir une taille fixe. Quelle que soit la quantité de données qu’on a à l’origine. Ensuite, la valeur de hachage est utilisée pour vérifier l’intégrité des données brutes.

Le second article utilise les propriétés de la cross-blockchain qui selon l’auteur possède les caractéristiques permettant de résoudre la problématique d’effacement des données et de droit à l’oubli [12] :

Le site web “cointribune” définit la cross-blockchain ou cross-chain de la façon suivante :

“Le terme cross-chain provient de l’anglais et peut être traduit approximativement en français par « inter-chaîne ». Dans l’univers des cryptomonnaies, c’est une expression qui est utilisée lorsqu’on parle de faire communiquer différentes blockchains entre elles.” [56]

“*Ensuite, Wanchain as a service (WaaS) est proposé pour s'assurer que les patients peuvent supprimer efficacement les informations de confidentialité stockées sur la blockchain en fonction de leur demande. Notre travail peut être résumé comme suit :*
- Nous proposons Wanchain en tant que service (WaaS). Nous adoptons Wanchain, qui est une technologie cross-blockchain utilisée pour résoudre le problème de fuite de confidentialité de la transmission des DSE dans différentes blockchains.
- Ce système permet aux patients de supprimer les données des DSE sur la blockchain. Il peut résoudre le problème que les données de la blockchain ne peuvent pas être supprimées et qu'une grande quantité de données est difficile à stocker sur la blockchain.
- Nous avons effectué une analyse de sécurité et une évaluation des performances pour prouver que notre solution est sûre, complète, raisonnable et réalisable.”

Concernant cette problématique, la CNIL recommande la solution suivante, utilisée par Arij Alfaidi & Edward Chow, décrite précédemment :

“*La CNIL constate qu’il est techniquement impossible de faire droit à la demande d’effacement de la personne concernée lorsque des données sont inscrites dans la Blockchain. Toutefois, lorsque la donnée inscrite sur la Blockchain est un engagement, une empreinte issue d’une fonction de hachage à clé ou un chiffré utilisant un algorithme et des clés conformes à l’état de l’art, le responsable de traitement peut rendre la donnée quasi inaccessible, et se rapprocher ainsi des effets d’un effacement de la donnée. En dehors du cas spécifique de certains engagements cryptographiques, ces solutions ne constituent pas*

un effacement de la donnée au sens strict dans la mesure où les données existeraient toujours sur la Blockchain. Néanmoins, la CNIL constate qu'elle permet de se rapprocher de l'exercice effectif de son droit à l'effacement pour la personne concernée." [50]

Cette partie nous a permis de constater que malgré tout, grâce à l'utilisation de la blockchain, la privacy des données est par défaut plus ou moins respectée. Les caractéristiques de la blockchain peuvent parfois être un frein au respect des recommandations légales, mais il existe des solutions pour pallier ces problématiques.

D) Type de blockchain utilisé

Comme on vient de le voir, pour répondre à certaines problématiques de privacy, il est recommandé de prioriser une blockchain de type permissioned dans le but de pouvoir donner des autorisations d'accès aux données de santé des patients.

Nous allons donc, dans cette section, nous intéresser aux différents types de blockchain existantes ainsi que le type de blockchain utilisé dans les publications. Il est intéressant de noter que contrairement aux notions de privacy, les chercheurs mentionnent explicitement le type de blockchain utilisé pour leur solution.

Il existe quatre types de blockchain [57] :

- *Public / permissionless* : La blockchain publique ne requiert aucune permission spécifique à l'entrée, ni au moment de réaliser une transaction. Les différents acteurs de la chaîne sont tous au même niveau, et tous les nœuds du réseau d'échange sont contrôlés par le réseau peer-to-peer. L'autre caractéristique essentielle de cette blockchain réside dans son caractère « open-source » : en effet, n'importe quelle personne disposant d'un bagage technique suffisant est en capacité de copier et de modifier le code du protocole selon son bon vouloir

- *Private / permissioned* : Pour se représenter ce qu'est un système de blockchain privée, on pourrait utiliser l'image d'un intranet privé, mettant à la disposition d'un nombre limité de participants un accès restreint à des contenus.

- *Hybrid : Les transactions et les enregistrements d'une blockchain hybride ne sont pas rendus publics, mais peuvent être vérifiés si nécessaire, par exemple en autorisant l'accès par un contrat intelligent. Les informations confidentielles sont conservées à l'intérieur du réseau, mais restent vérifiables. Même si une entité privée peut posséder la blockchain hybride, elle ne peut pas modifier les transactions.*
  Lorsqu'un utilisateur rejoint une blockchain hybride, il a un accès complet au réseau. L'identité de l'utilisateur est protégée des autres utilisateurs, sauf s'il effectue une transaction. Dans ce cas, son identité est révélée à l'autre partie.

- *Consortium :* la blockchain de consortium, également appelée blockchain fédérée, est similaire à une blockchain hybride en ce sens qu'elle possède des caractéristiques de blockchain privée et publique. Mais elle est différente dans la mesure où plusieurs membres de l'organisation collaborent sur un réseau décentralisé. Essentiellement, une blockchain de consortium est une blockchain privée dont l'accès est limité à un groupe particulier, ce qui élimine les risques liés au contrôle du réseau par une seule entité sur une blockchain privée.
  Dans une blockchain de consortium, les procédures de consensus sont contrôlées par des nœuds prédéfinis. Elle dispose d'un nœud validateur qui initie, reçoit et valide les transactions. Les nœuds membres peuvent recevoir ou initier des transactions.

Le type de blockchain le plus utilisé dans notre base de publication est de loin le type private / permissioned :

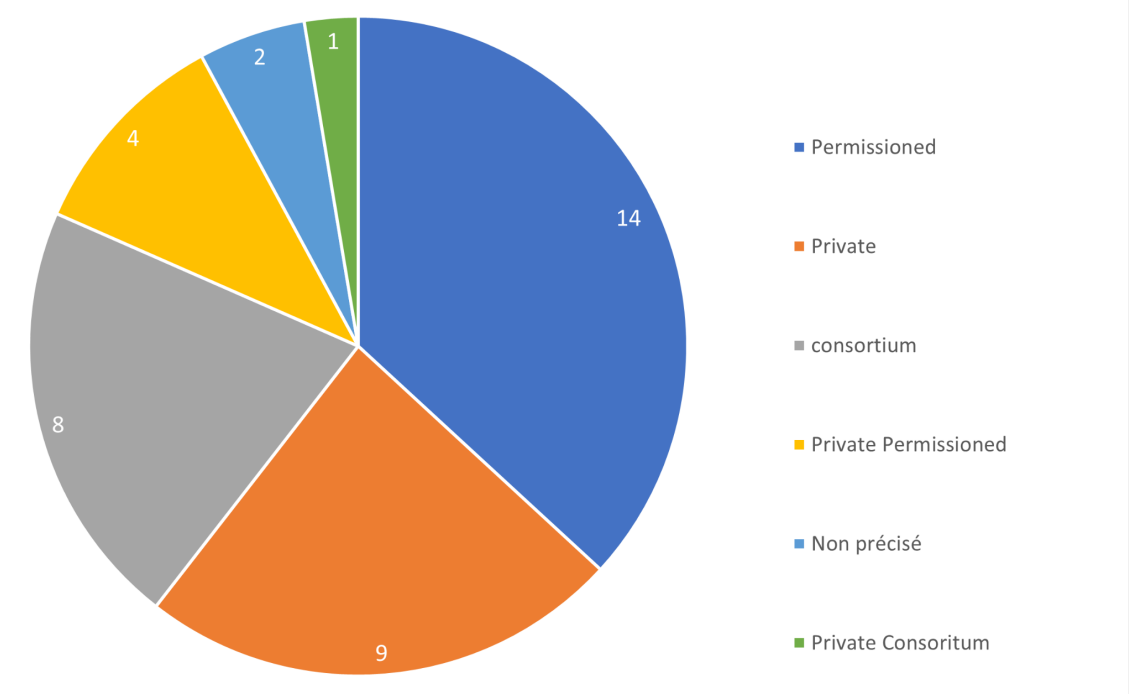


Figure 9 : Type de blockchain utilisé

En effet, ce type est utilisé par quasiment 26 soit 70% des publications (Permissioned + Private + Private Permissioned + Private consortium)

Les auteurs comprennent donc que dans le cadre du traitement de données de santé, il est préférable voire indispensable d'utiliser une blockchain permissioned ou du même type. C'est d'ailleurs pour cela que nous pouvons remarquer qu'aucune architecture présentée dans les publications utilisent une blockchain de type publique. Cela rendrait les données accessibles à tous et aucun contrôle ne pourrait être effectué

E) Classification de l'approche de l'article et niveau de conférence des articles

Dans l'optique d'analyser et évaluer le type d'article que nous traitons dans le cadre de notre revue systématique de littérature, nous avons choisi de classer les articles selon leur approche pour la littérature. Nous avons utilisé la base apportée par Roel Wieringa, Neil Maiden et Nancy Mead Colette Rolland [58]. Les auteurs font une méthodologie pour la classification des articles, selon plusieurs critères, qui s'inscrivent dans l'une des six catégories suivantes :

- **Evaluation research**: Le but de la recherche est d'aboutir à l'établissement d'une nouvelle connaissance des relations causales entre les phénomènes ou à de nouvelles connaissances logiques entre les propositions.

- **Proposal of solution :** Propose une solution technique sur un sujet et argumente sur la pertinence de sa solution. Dans cette catégorie, la solution n'a pas été validée par les pairs. La technique doit être nouvelle, ou au moins une amélioration significative d'une technique existante.

- **Validation research**: Étude des propriétés d'une proposition de solution qui n'a pas encore été mise en œuvre dans la pratique de l'Ingénierie des exigences. L'enquête utilise une configuration de recherche approfondie et méthodologiquement solide. Les méthodes de recherche possibles sont les expériences, la simulation, le prototypage, l'analyse mathématique, la preuve mathématique des propriétés, etc.

- **Philosophical papers**: Apporte une nouvelle manière de visualiser, d'analyser, de conceptualiser une situation.

- **Opinion papers**: Article contenant l'opinion des auteurs sur un sujet, la façon de l'aborder, les bonnes pratiques qui pourraient être appliquées selon lui…

- **Personal experience papers**: Dans ces articles, l'accent est mis sur le quoi et non sur pourquoi. Le document doit contenir une liste des leçons tirées par l'auteur de son expérience. Les articles de cette catégorie proviendront souvent de praticiens de l'industrie ou de chercheurs qui ont utilisé leurs outils dans la pratique, et l'expérience sera rapportée sans discussion des méthodes de recherche.

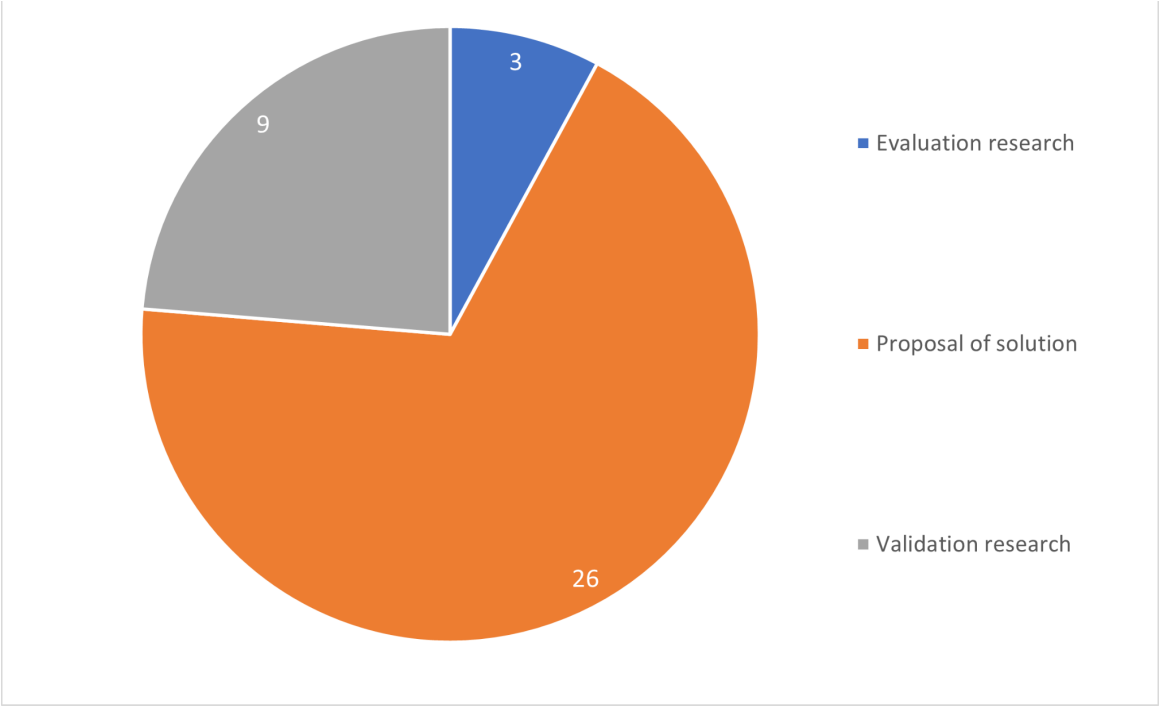


Figure 10 : Classification de l'approche des articles

La figure ci-dessus nous montre que 70% de nos articles sont des proposal of solution. Cela est totalement cohérent avec le sujet de notre problématique qui analyse les solutions utilisant la blockchain pour l'échange de données de santé.

Les approches de type philosophical papers, opinion papers et personal experience papers, ne sont pas présentes dans notre base d'article. Cela s'explique, car ces approches font un recul sur une technologie ou autre, mais la blockchain est une technologie bien trop récente pour pouvoir faire cela actuellement.

Nous avons utilisé les outils CORE portal et conférence. CORE établit un classement qui évalue les principales conférences dans lesquelles sont publiés les articles et journaux dans le domaine de l'informatique.

Les conférences sont classées dans l'une des catégories suivantes :

- A* : Conférence de très haut niveau, conférence la plus respectée dans le domaine concerné
- A : Conférence d'excellente qualité et très respecté dans un domaine disciplinaire
- B : Bonne à très bonne conférence, et bien considéré dans un domaine disciplinaire
- C : Conférence autres étant classés qui répondent aux normes de base

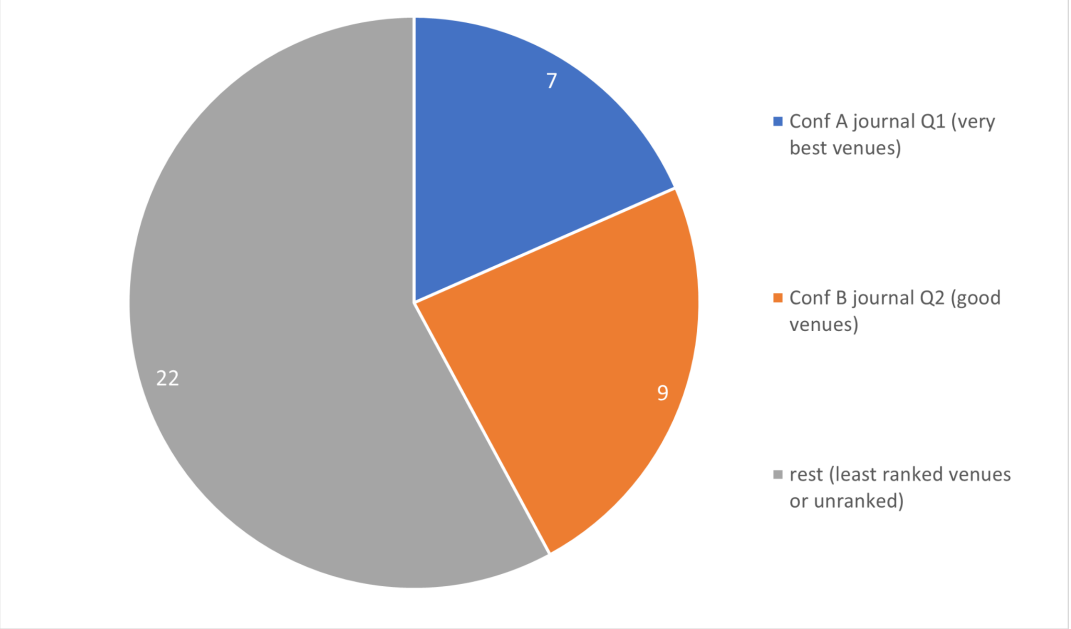


Figure 11 : Qualité de conférence des articles

Nous considérons que les articles faisant partie d'une conférence avec rang A+ ou A font partie d'une bonne conférence, les articles faisant partie d'une conférence notée B font partie d'une bonne conférence et pour le reste faisant partie d'une conférence de moins bonne facture.

Concernant notre base d'articles, nous constatons que 58% des conférences ne font pas partie d'une conférence de rang A+ / A / B.
Parmi les articles faisant partie d'une bonne conférence, nous retrouvons l'unique article définissant et reliant la privacy à des notions légales au sein de sa solution. Il s'agit de l'article Hyperledger fabric blockchain : Secure and efficient solution for electronic health records [30] avec un rang A.

Nous retrouvons également nos deux articles issus du snowballing faisant partie d'une conférence de rang B.

Nous constatons ici que trois articles parmi les quatre font parties de bonne conférence. Cela laisse entendre que le niveau de la conférence semble être un bon indicateur sur le niveau de preuve fourni par les articles pour appuyer et documenter sur la qualité de leur solution.

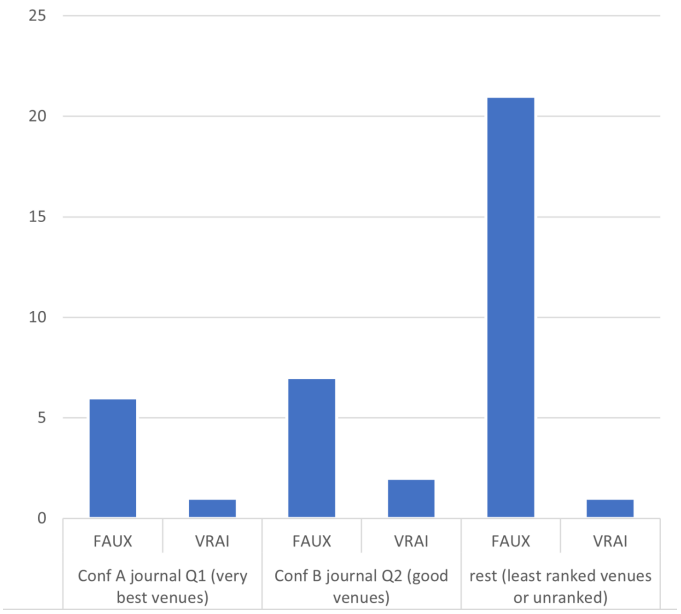


Figure 12 : Association des articles définissant la privacy avec le niveau de la conférence

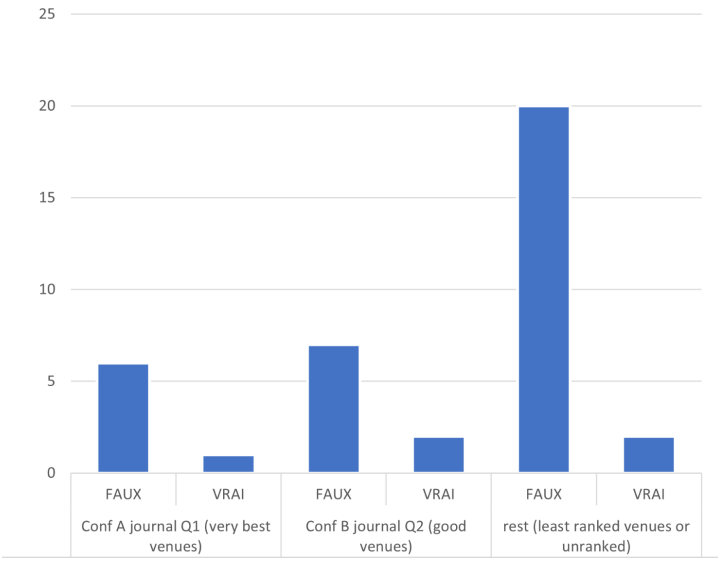


Figure 13 : Association des articles reliant la privacy à des notions légales avec le niveau de la conférence

Nous constatons avec les deux figures ci-dessus qu'entre 20 et 30% des articles définissants et/ou la reliant à des notions légales sont de conférences d'un rang A+, A ou B. C'est deux à trois fois moins pour les articles de rang inférieur.

2) Analyse qualitative

A. Linddun GO : modélisation des menaces à la vie privée

Dans cette partie, nous allons étudier un article plus en profondeur et voir sa conformité à la privacy grâce à l'outil Linddun go.

Au sens le plus général, la modélisation des menaces (threat modeling) correspond à une analyse systématique des risques, des attaques, des vulnérabilités potentielles, avec l'objectif de mettre en place des protections efficaces contre les menaces identifiées.

Voici quelques définitions de ce sujet dans la littérature :

- Uzunov et Fernandez (2014) [59], "la modélisation des menaces est un processus qui peut être utilisé pour analyser les attaques ou les menaces potentielles, et peut également être soutenu par des bibliothèques de menaces ou des taxonomies d'attaques".

- Bedi et al., (2013) [60], la modélisation des menaces "fournit une manière structurée de concevoir des logiciels sécurisés, ce qui implique de comprendre l'objectif d'un adversaire pour attaquer un système en fonction des actifs d'intérêt du système ".

- Baquero et al., (2015) [61], “Bien que centrée sur le développement d'applications, la modélisation des menaces est la technique qui aide les ingénieurs logiciels à identifier et à documenter les menaces de sécurité potentielles associées à un produit logiciel, fournissant aux équipes de développement un moyen systématique de découvrir les forces et les faiblesses de leurs applications logicielles ".

C’est une activité essentielle pour identifier, détecter des failles et pouvoir mettre en place des solutions pendant la phase de conception de l’application avant même de passer à la phase de développement du système ou du logiciel. La modélisation des menaces est surtout utilisée pour identifier des menaces à la sécurité.

Linddun GO est un outil permettant de faire de la modélisation des menaces pour la vie privée ou également appelée “Privacy by Design”. Le Privacy By Design est un concept faisant partie intégrante du Règlement Général pour la Protection des Données (RGPD).

Le terme "Privacy by Design" ne signifie rien d'autre que la protection des données dès la conception technologique. L'idée sous-jacente est que la protection des données dans les procédures de traitement des données est mieux respectée lorsqu'elle est déjà intégrée dans la technologie lors de sa création.

L’objectif du privacy by design est d’impliquer la protection de données personnelles comme une partie intégrante des réflexions en amont de la conception des processus métiers. En

effet, dès la réflexion sur la modélisation d'un processus, le respect de la vie privée doit faire partie des priorités absolue.

Voici la description de Linddun GO selon leur site internet "Linddun GO est conçu pour vous donner un démarrage rapide dans la modélisation des menaces à la vie privée. Il s'agit d'une approche de modélisation des menaces structurée selon les catégories de menaces Linddun. Il vise à fournir un support structuré, mais léger pour la modélisation des menaces.

Le support de connaissances fourni est divisé en 7 catégories de menaces, qui sont encapsulées dans l'acronyme LINDDUN.

1. **Liabilité** : Un adversaire est capable de lier deux éléments d'intérêt sans connaître l'identité de la ou des personnes concernées.

2. **Identifiabilité** : Un adversaire est capable d'identifier une personne concernée parmi un ensemble de personnes concernées par le biais d'un élément d'intérêt.

3. **Non-répudiation :** La personne concernée n'est pas en mesure de refuser une réclamation (par exemple, avoir effectué une action ou envoyé une demande).

4. **Détectabilité** : Un adversaire est capable de distinguer si un élément d'intérêt concernant une personne concernée existe ou non, indépendamment de sa capacité à lire le contenu lui-même.

5. **Divulgation d'informations :** Un adversaire est capable d'apprendre le contenu d'un élément d'intérêt au sujet d'une personne concernée.

6. **Inconscience** (Unawareness) : La personne concernée n'est pas au courant des activités de collecte, de traitement, de stockage ou de partage (et des finalités correspondantes) des données personnelles de la personne concernée.

7. **Non-conformité** : Le traitement, le stockage ou la manipulation des données personnelles n'est pas conforme à la législation, à la réglementation et/ou à la politique. " [62].

Fonctionnement de l’outil

Contrairement à Linddun, Linddun GO a la particularité d’être accessible à tous. Il fonctionne sous forme de jeu de cartes pour identifier de potentielles menaces à la vie privée et ainsi évaluer la privacy d’une architecture logicielle.

Pour y “jouer” et analyser la privacy d’une architecture, Linddun GO recommande les trois choses suivantes :

1. Être entre 2 à 5 personnes

2. Avoir un modèle intégrant les éléments correspondants aux types de points d'accès utilisés par LINDDUN GO

3. Piocher à tour de rôle une carte et identifier une menace applicable pertinente.
    a. Q1 (pourrait-il être fait ?) aide à déterminer si les conditions préalables de la menace sont remplies et si la menace pourrait se produire.
    b. Q2 (serait-ce un problème ?) permet d'évaluer si la menace est réellement applicable.

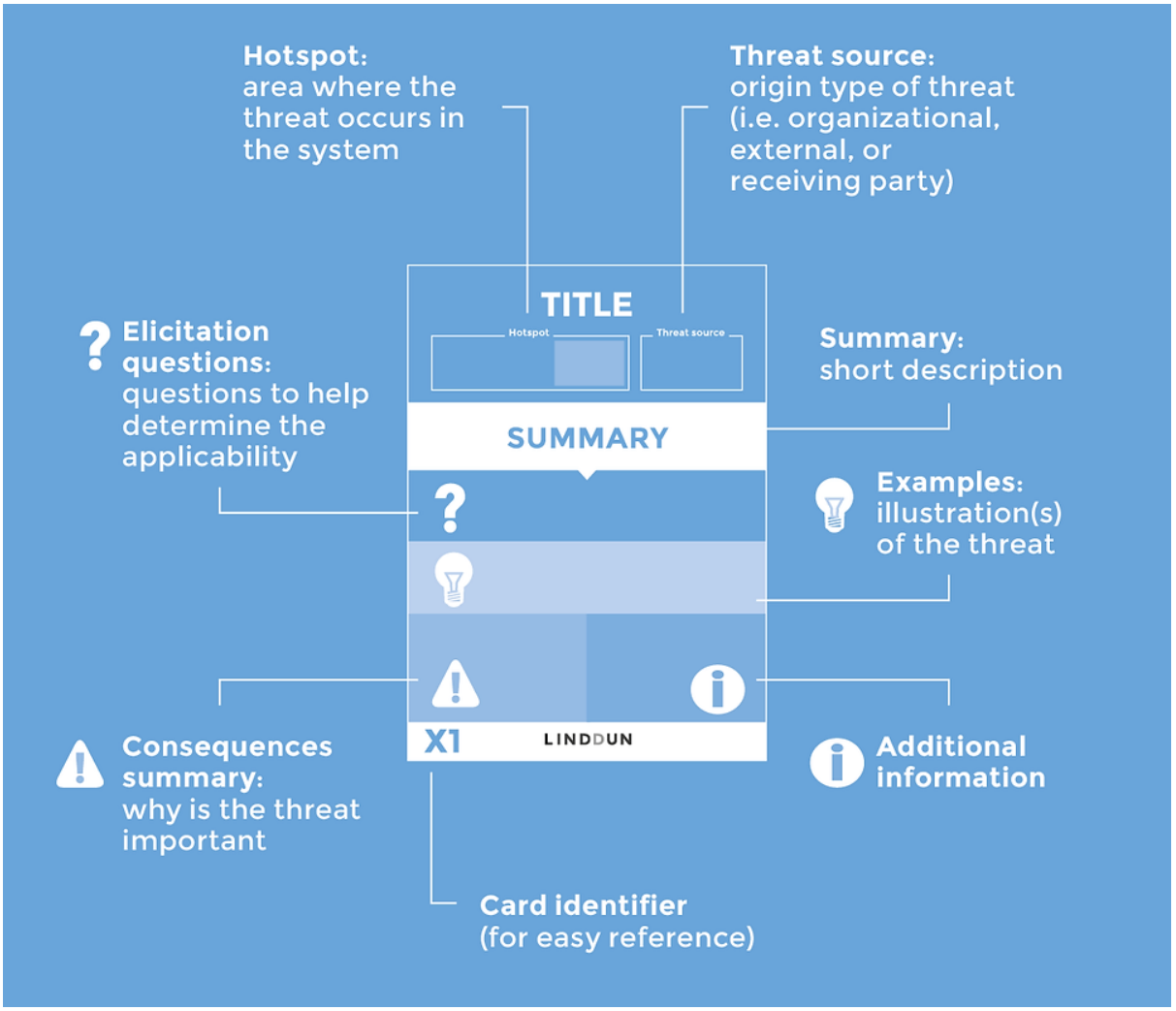
Si on peut répondre "oui" aux deux questions pour un hotspot spécifique, nous sommes face à une menace potentielle. Super ! N'oubliez pas de le documenter.

Il y a différents modes de jeu pour cet outil, nous avons choisi d’utiliser le mode “Time-boxed” pour une utilisation rapide, mais structurée sur un article.

Ce mode de jeu sert à limiter le temps de l'exercice (ou limitez le nombre de cartes) et effectuez plusieurs sessions de modélisation des menaces

Les cartes Linddun GO sont répartis en plusieurs sections :

Les cartes de type de menace décrivent les menaces potentielles qui peuvent survenir.
Chaque carte est structurée comme suit :e



Annexe 1 : Modèle de carte Linddun

Chaque carte de type de menace suit le même modèle [71], comme l'exemple présenté ci-dessus.

- Titre : Titre du type de menace.
- Hotspots : La zone où la menace se produit dans le système
- Source de la menace. Type d'origine de la menace (c.-à-d., organisationnel, externe au système ou la partie réceptrice de l'interaction)
- Résumé : Brève description du type de menace
- Questions d'élicitation : Deux questions pour aider à déterminer l'applicabilité du type de menace.
  - La première question détermine principalement si les conditions préalables sont remplies
  - La seconde question aide à évaluer l'applicabilité elle-même, l'applicabilité elle-même.
- Exemples : Illustration(s) du type de menace.
- Conséquences/impact : Justification de l'importance de la menace.

- ❖ Informations complémentaires : Complète la description avec des informations supplémentaires sur le type de menace.
- ❖ Identifiant de la fiche : Identifiant pour une référence facile.
- ❖ Catégorie LINDDUN : Mise en évidence de la catégorie de menace, catégorie de menace correspondante dans l'acronyme LINDDUN.

Il y a six types de cartes pour analyser différents types de menaces à la privacy :

- Linkability
- Identifiability
- Non-repudiation
- Detectability
- Unawareness
- Non-compliance

Choix de la publication

Nous allons voir dans cette section le fonctionnement de l'outil en analysant une publication.

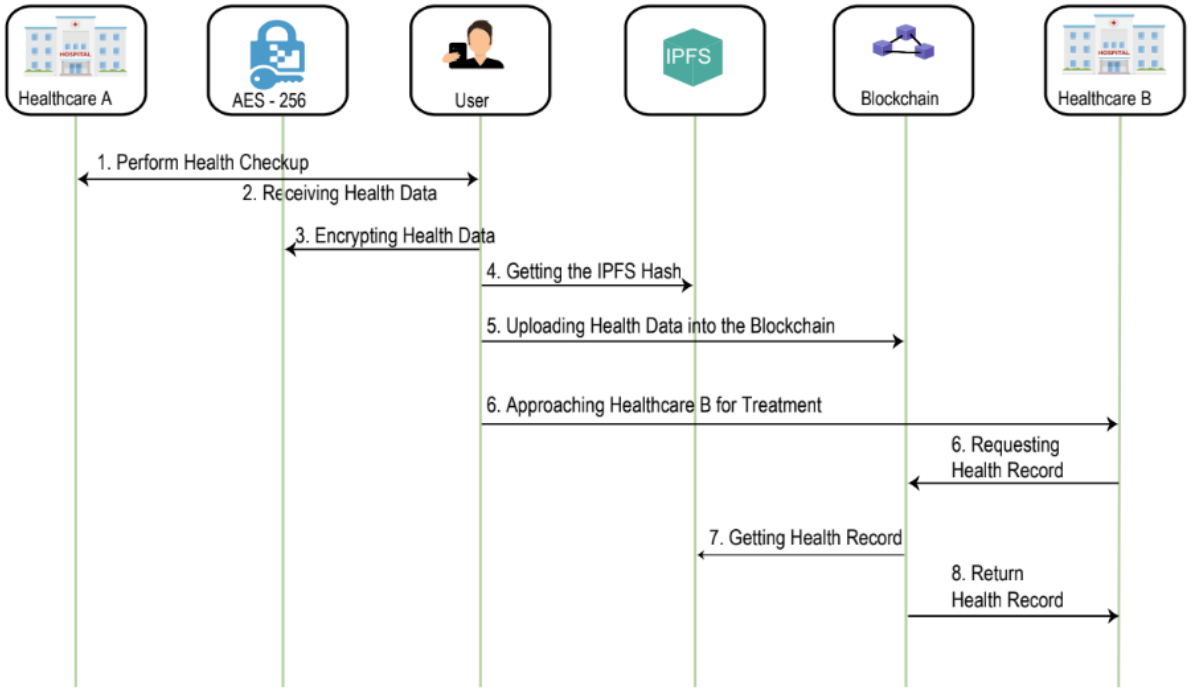
Pour ce faire, nous avons choisi l'article suivant "A Combined Framework of InterPlanetary File System and Blockchain to Securely Manage Electronic Medical Records" [2].

Nous avons choisi cet article, car il représente bien l'ensemble des articles faisant parties de notre base de publications. La solution proposée utilise une blockchain de type private / permissoned, la privacy n'est pas défini par les auteurs et aucune réglementation n'est mentionnée concernant les aspects à respecter pour se mettre en conformité avec la vie privée.

Analyser les menaces pour la privacy sur cette publication est une bonne façon de voir si les publications ne s'intéressant pas à la privacy au premier plan peuvent quand même être en conformité avec celle-ci. Nous avons déjà constaté via l'analyse quantitative que les publications faisant références à des lois concernant la vie privée sont plus attentives à l'utilisation des données personnelles et le respect des contraintes réglementaires au sein de leur solution.

Il était donc préférable de s'intéresser à un article ne mettant pas l'accent sur la privacy sachant de plus que ce type d'article est largement plus représentatif de notre base de publication.

De plus, il présente un diagramme de séquence, présentant l'architecture du workflow pour le stockage et l'échange de la donnée, de la solution proposée.



Annexe 2 : Proposition de flux de travail pour le stockage et la distribution sécurisés des DME sur la blockchain. [2]

L'outil recommande l'utilisation d'un diagramme flux de données (DFD) mais mentionne l'usage d'un diagramme de séquence le cas échéant. N'ayant pas trouvé d'article utilisant un diagramme flux de données pour représenter l'échange de donnée à travers la solution, nous avons donc opté pour celui-ci qui propose un diagramme de séquence du flux de travail pour le stockage et la distribution des données.

Utilisation de l'outil

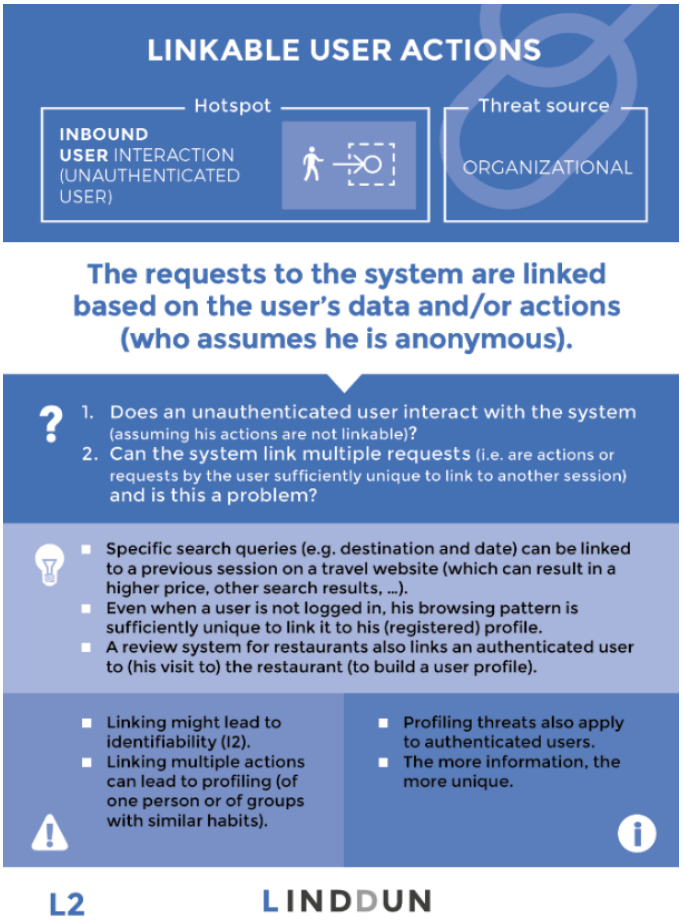
Nous avons utilisé l'ensemble des six types de cartes proposées par l'outil [63] :

- 2 cartes de types Linkability
- 1 cartes de types Identifiability
- 1 carte de types Non-repudiation
- 1 carte de type Detectability
- 2 cartes de types Unawareness
- 2 cartes de types Non-compliance

Nous allons ci-dessous voir les cartes et analyses effectuées grâce à l'outil.

Cartes de types Linkability

Ces cartes sont utiles pour être capable de distinguer suffisamment si deux éléments d'intérêt sont liés ou non, même sans connaître l'identité réelle du sujet des éléments d'intérêt pouvant être liés.



Annexe 3 : Carte de type Linkability 1 [73]

"Les demandes adressées au système sont liées sur la base des données et/ou des actions de l'utilisateur (qui suppose qu'il est anonyme)."

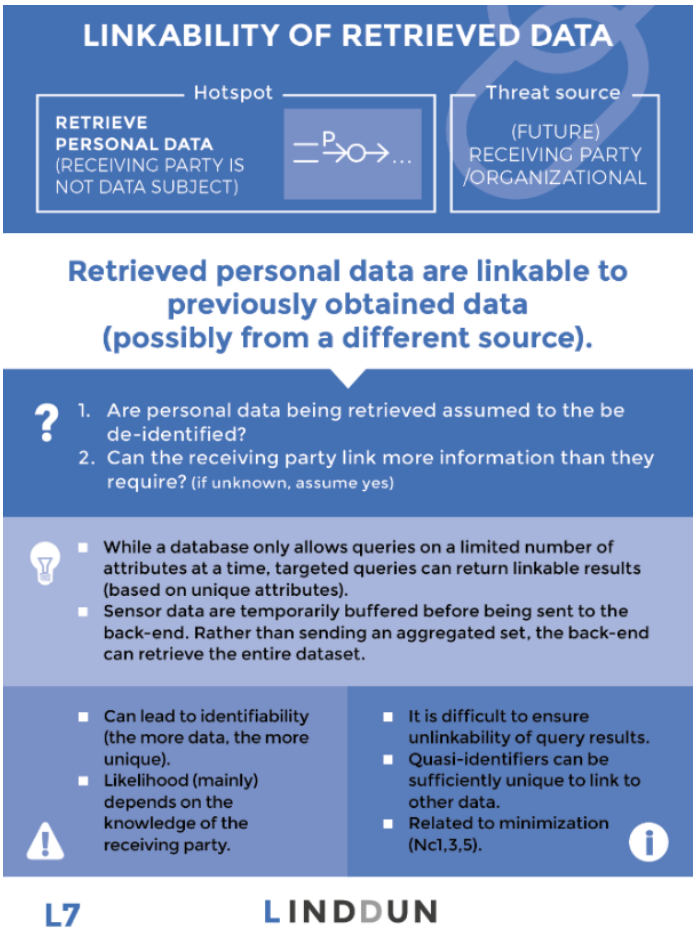
Oui, les demandes adressées au système sont liées sur la base des données car les données sont stockées sur une base de données externe nommée IPFS et leurs hash est stocké sur la blockchain.

"1. Un utilisateur non authentifié interagit-il avec le système (en supposant que ses actions ne puissent pas être liées) ?"

Un utilisateur non authentifié ne peut interagir avec le système parce que le type de blockchain utilisé nécessite une autorisation pour accéder à son contenu. De plus, l'utilisateur a besoin de l'accord du propriétaire des données.
"Dans le système proposé, si un utilisateur souhaite accéder aux rapports médicaux d'un patient, il doit obtenir la clé privée du propriétaire des données, car celles-ci sont cryptées à l'aide de l'algorithme AES-256." [2]

"2. Le système peut-il lier plusieurs demandes (c'est-à-dire que les actions ou les demandes de l'utilisateur sont-elles suffisamment uniques pour être liées à une autre session) et cela pose-t-il un problème ?"

Les informations données par la publication ne nous permettent pas de répondre à la question



Annexe 4 : Carte de type Linkability 2 [73]

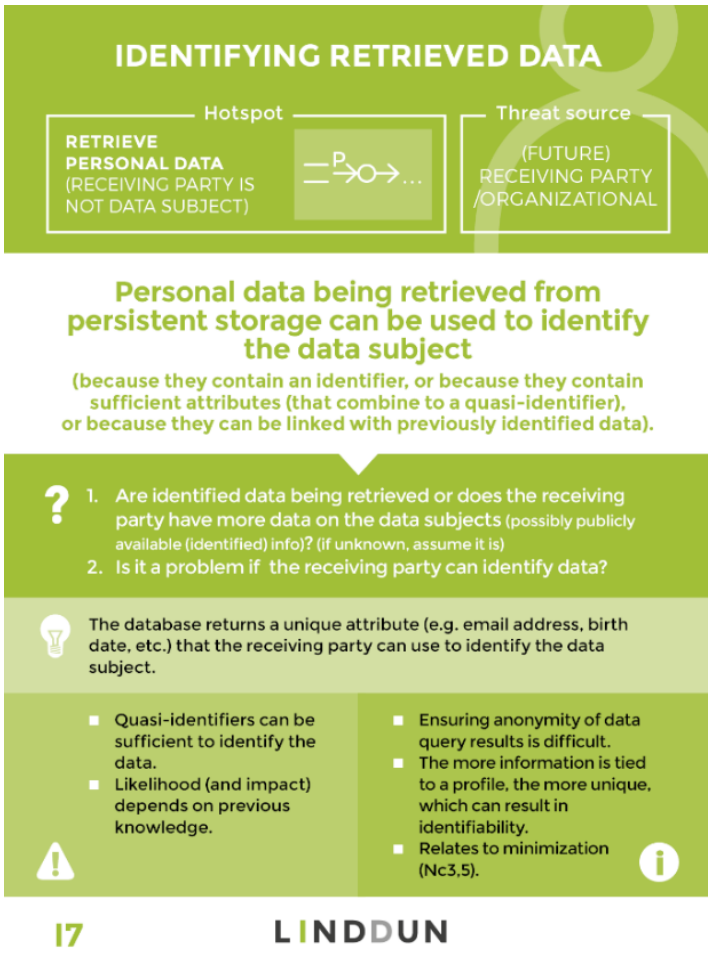
"Les données à caractère personnel récupérées peuvent être reliées à des données obtenues précédemment (éventuellement d'une source différente)."

Dans le cadre des données de santé, d'autre source d'informations peuvent être reliée aux données récupérées, que ce soit la même source ou une source externe.

“1. Les données personnelles extraites sont-elles supposées être dépersonnalisées ?”
Étant donné que nous utilisons des données médicales, il parait que toutes les infos extraites sont personnelles.
“2. La partie réceptrice peut-elle associer plus d'informations qu'elle en a besoin ? (si l'on ne sait pas, on suppose que oui)”

Oui par exemple dans le cadre d’une extraction de DSE la personne aura tous les antécédents de santé du patient dont potentiellement des informations sans intérêt pour le cas actuel

Cartes de types Identifiability



Annexe 5 : Carte de type Identifiability [73]

“Les données à caractère personnel extraites d'un stockage permanent peuvent être utilisées pour identifier la personne concernée (parce qu'elles contiennent un identifiant, ou parce qu'elles contiennent suffisamment d'attributs (qui se combinent à un quasi-identifiant ou parce qu'elles peuvent être liées à des données précédemment identifiées).”

Oui, elles peuvent être utilisées pour identifier la personne concernée.
“1. Les données identifiées sont-elles récupérées ou la partie destinataire dispose-t-elle de plus de données sur les personnes concernées (éventuellement des informations (identifiées) accessibles au public) ? (si vous ne le savez pas, supposez que c'est le cas)”

Il semble que non. La partie destinataire ne dispose pas de plus de donnée sur les personnes concernées, mais les données sont encryptées

“2. Est-ce un problème si la partie réceptrice peut identifier les données ?”

Non, car la partie réceptrice est un autre centre de santé qui demande les infos du patient

Cartes de types Non-repudiation

Annexe 6 : Carte de type Non-repudiation [73]

“La personne ne peut pas nier s'être authentifiée à un service”

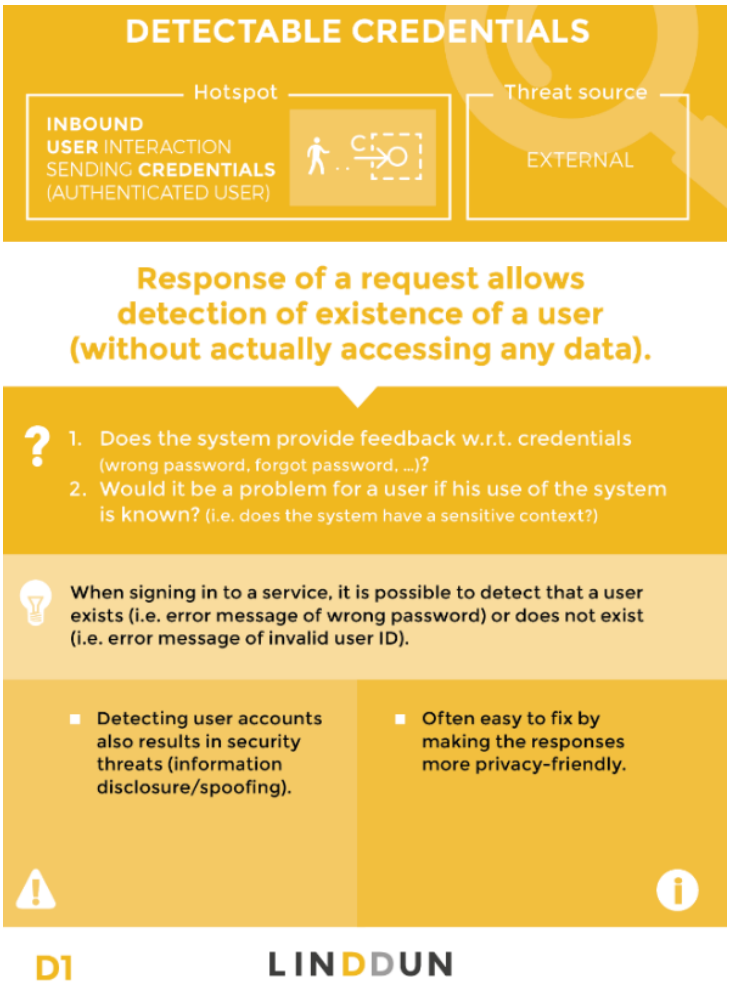
Elle ne peut le nier puisque sur la blockchain il y a une traçabilité de toutes les actions

“1. Le système exige-t-il des informations d'identification ?”

Oui, par exemple, il requiert l’utilisation de la clé privée de l’utilisateur : “Enfin, l'utilisateur peut partager sa clé privée pour décrypter le document au personnel autorisé.” [2]
“2. Cela pose-t-il un problème si un utilisateur peut être lié au système (lorsque les informations d'identification ou les journaux d'accès sont divulgués) ? (Il est peu probable que cette menace soit applicable. Seulement lorsque l'utilisation du système est considérée comme "sensible")”

Non, le secteur d’activité et l’utilisation du système n’est pas considérée comme sensible

Cartes de types Detectability



Annexe 7 : Carte de type Detectability [73]

“La réponse à une requête permet de détecter l'existence d'un utilisateur (sans accéder réellement à des données).”

Oui, car chaque requête permet de récupérer les données d’un user (patient) précis.

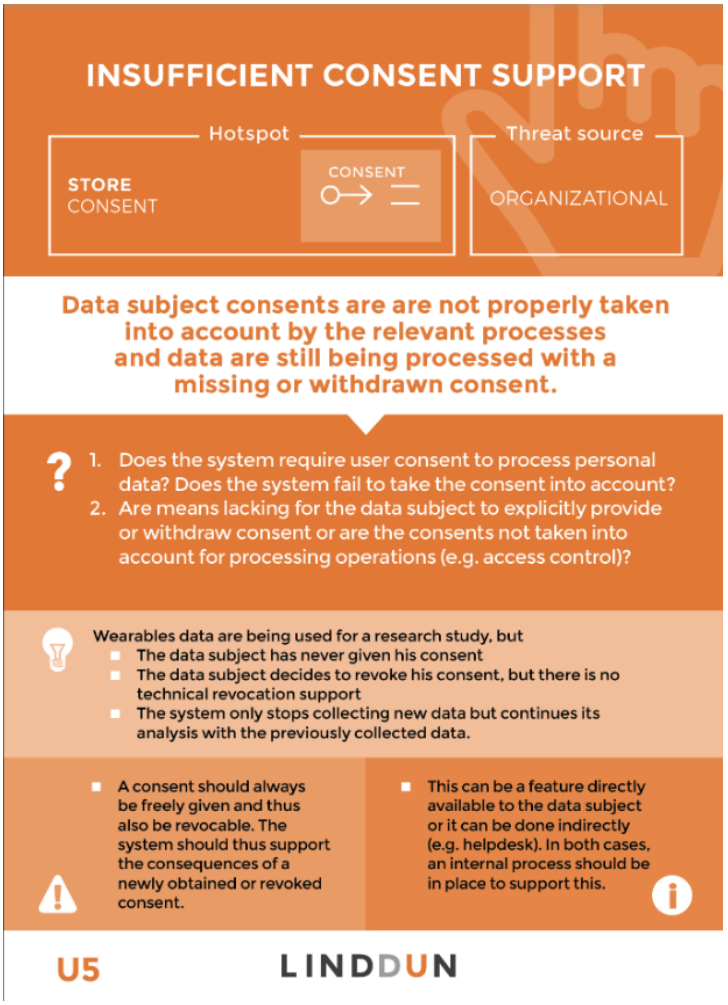
“1. Le système fournit-il un retour d'information sur les informations d'identification (mauvais mot de passe, mot de passe oublié, etc.) ? (mauvais mot de passe, mot de passe oublié) ?”

Les informations fournies par la publication ne nous permet de répondre à cette question.

2. Est-ce que ce serait un problème pour un utilisateur si son utilisation du système est connue ? (c.-à-d. le système a-t-il un contexte sensible ?)

Non, cela n’est pas un problème, car le système n’a pas un contexte sensible.

Cartes de types Unawareness



Annexe 8 : Carte de type Unawareness 1 [73]

“Les consentements des personnes concernées ne sont pas correctement pris en compte par les processus pertinents et les données sont toujours traitées en l'absence de consentement ou avec un consentement retiré.”

Les personnes concernées doivent donner leurs permissions pour décrypter les données, cela peut s'apparenter à un consentement.
"1. Le système exige-t-il le consentement de l'utilisateur pour traiter les données personnelles ? Le système ne prend-il pas en compte le consentement ?"

Oui, le système impose le consentement de l'utilisateur pour traiter les données personnelles, le patient doit interagir pour stocker le DSE sur le réseau IPFS

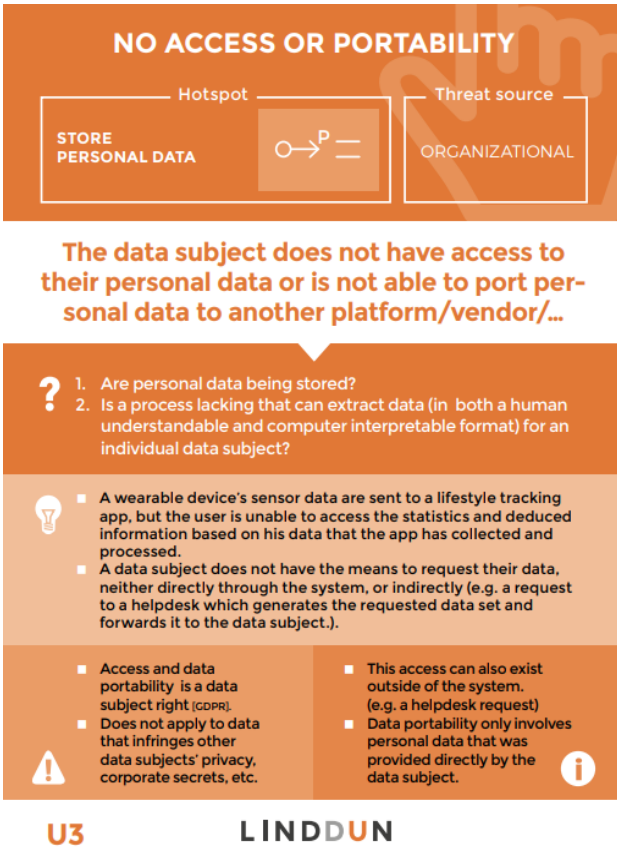
"*Obtenir le hachage de l'IPFS : Après avoir reçu le dossier médical électronique de l'établissement de santé A, le patient le chiffre d'abord, puis le télécharge sur le réseau IPFS pour recevoir une clé de hachage SHA256.*

Ajout du DME sur la blockchain : Au cours de l'étape de téléchargement, le patient doit fournir le nom de fichier du DME afin de pouvoir, à l'aide de ce nom, interroger et extraire le DME de la blockchain si nécessaire (figure 4c). En plus du nom du DME, le patient fournit également son nom et la clé de hachage unique reçue de l'IPFS." [2]

"2. Les moyens permettant à la personne concernée de donner ou de retirer explicitement son consentement font-ils défaut ou les consentements ne sont-ils pas pris en compte pour les opérations de traitement (par exemple, le contrôle d'accès) ?"

Le document ne parle pas de consentement explicite, mais il est fait mention d'une forme de contrôle d'accès :

"Analyse de la sécurité : Dans le système proposé, si un utilisateur souhaite accéder aux rapports médicaux d'un patient, il doit obtenir la clé privée du propriétaire des données"



Annexe 9 : Carte de type Unawareness 2 [73]

"La personne concernée n'a pas accès à ses données personnelles ou n'est pas en mesure de les transférer vers une autre plateforme/fournisseur/... ?"

Non, la personne concernée peut accéder à ses données personnelles et il semble qu'il peut les transférer vers une autre plateforme.
"Il peut facilement partager le DME de la blockchain avec d'autres autorités de santé, comme les pharmacies, les médecins ou les agences d'immigration." [2]

"1. Les données personnelles sont-elles stockées ?"

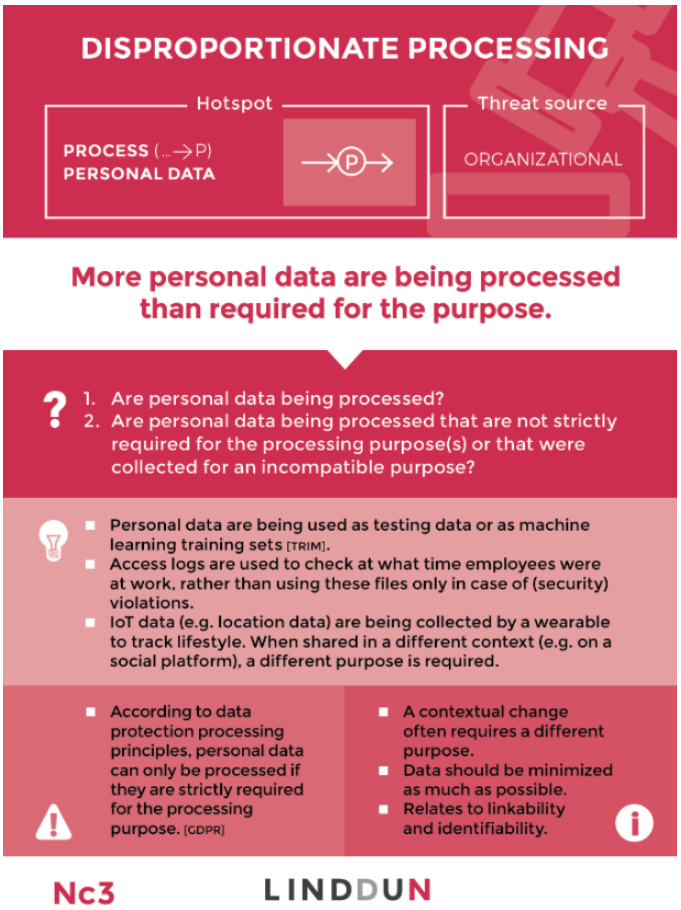
Oui sur une IPFS, base de donnée stockée dans le cloud

"2. Il manque un processus permettant d'extraire des données (dans un format compréhensible par l'homme et interprétable par l'ordinateur) pour une personne concernée ?"

Non, il ne manque pas de processus dans ce cas. L'extraction se passe de la maniere suivante :

➔ L'entité récupère les données de la base de donnée IPFS par l'intermédiaire de la blockchain

Cartes de types Non-compliance



Annexe 10 : Carte de type Non-compliance 1 [73]

“Le nombre de données à caractère personnel traitées est supérieur à ce que requiert la finalité.”

Non, car les données récupérées sont des informations de santé sur le patient qui peuvent être utile par la suite pour les médecins

“1. Des données à caractère personnel sont-elles traitées ?”
Oui car ce sont des données de santé donc très sensibles

“2. Des données à caractère personnel sont-elles traitées alors qu'elles ne sont pas strictement requises pour la ou les finalités du traitement ou qu'elles ont été collectées pour une finalité incompatible ?”

Non, parce que la finalité de recueillir l’ensemble des informations peut s’avérer nécessaire au traitement du patient

Annexe 11 : Carte de type Non-compliance 2 [73]

“Il n'existe pas de motif légitime pour la collecte ou le traitement ultérieur et le stockage des données à caractère personnel.”

Non, il existe bien un motif légitime, le traitement et le suivi des données de santé du patient

“1. Des données à caractère personnel sont-elles traitées dans le système ?”

Oui, des données à caractère personnel sont traitées dans le système

“2. La personne concernée ne consent-elle pas au traitement et n'y a-t-il pas d'autre motif légitime ?”
Elle ne consent pas forcément de manière explicite, mais dans le cadre des données de santé, le consentement n’est pas impératif.

Conclusion sur l'utilisation de Linddun GO

La privacy prend de plus en plus d'ampleur, elle devient règlementée et les applications développées doivent se mettre en conformité pour respecter les données personnelles.

Le cadre d'ingénierie de confidentialité Linddun GO nous permet d'analyser les menaces de confidentialité en fonction des différentes catégories de menaces identifiées par l'outil. Les questions proposées par ce "jeu" nous permet d'analyser si la solution est en adéquation avec la vie privée des utilisateurs.

Ces questions nous invite à analyser le stockage, l'échange, la réception et l'envoi de la donnée avec un objectif : identifier des menaces à la vie privée et pouvoir ainsi produire une solution à ces menaces.

Pour rappel, si l'on peut répondre "oui" aux deux questions posées sur la carte de menace proposée par l'outil, on peut considérer que la solution contient une faille potentielle.

Il y a qu'une seule carte dans ce cas (Annexe 4 : Carte de type Linkability 2). Dans certains cas, nous n'avons pas pu répondre à la question par manque d'informations dans l'article.

Pour autant, cela signifie que nous n'avons pas détecté d'autres menaces à la privacy pour la publication analyser.

L'utilisation de Linddun GO sur une publication de notre base nous permet de conclure que la solution ne semble pas contenir de failles majeures contre la privacy.

Afin d'approfondir, il serait approprié de refaire cette démarche avec des experts en privacy pour analyser de la meilleure façon les failles que pourrait contenir la solution.

B. Checklist des aspects à prendre en compte quand on traite la privacy

Il est de ce fait important d'examiner la contribution à la privacy pour la solution que l'on propose. Pour cela, nous avons étudié les choses à regarder pour optimiser sa conformité avec la privacy afin de proposer une checklist pour visualiser sa bonnes conformités avec la privacy. D'autant plus qu'il est bon de rappeler que les chercheurs ont la responsabilité d'examiner et de respecter les exigences légales dans le pays où ils exercent, mais également dans les pays où leurs solutions seront mises en place.

Cette checklist a pour objectif de donner au chercheur les aspects à vérifier et identifier dans le cadre de la privacy et de la protection des données. Elle a été établie depuis les informations récoltées sur la Commission nationale de l'informatique et des libertés (CNIL) et l'Organisation de coopération et de développement économiques (OCDE).

Les points à examiner sont donc les suivants :

- **Identifier le(s) responsable(s) de traitement**
  L'identité et les responsabilités de gestion des données personnelles incombées à la personne s'en occupant doivent être claires et énumérées.

- **Identifier le(s) sous-traitant(s) des données :**
  S'il est fait usage d'externalisation ou sous traitance, il faut que les données soient soumises aux mêmes critères stricts par les sous-traitants pour s'assurer de privacy des données. Ces sous-traitants doivent être identifiés et leurs responsabilités clairement établies.

- **Récolter uniquement les données nécessaires :**
  Pendant la conception de votre solution, il faut se questionner et s'assurer que vous ne récolterez seulement les données nécessaires au fonctionnement de votre solution. Les indications issues des réglementations requièrent de ne récolter que les informations utiles au besoin de la solution. Il est donc important de s'assurer de la légitimité du recueil de ses données.

- **Obtenir le consentement des personnes concernées par la collecte de donnée :**
  Dans le cadre d'appréciation médicale, l'obtention du consentement des personnes concernées n'est pas exigée.
  Dans le cas contraire, le consentement est soumis aux quatre critères suivants :
  - Libre : ne doit pas être contraint, la personne doit être volontaire
  - Spécifique : doit être liée à un ou des objectifs clairs et identifiés
  - Éclairé : la personne doit être en pleine connaissance de la finalité du traitement et de la nécessité du recueil de ses informations. Dans le cadre de la collecte de données, vous vous devez d'être claires sur la finalité et l'objectif pour lesquelles les données sont collectées.

    - Univoque : doit être obtenue par une déclaration ou autres actes positifs de la personne.

- **S’assurer de l’intégrité, la sécurité et le stockage des données :**
  Mettre en place des procédures pour s’assurer que les données soient exactes et à jour. Cette partie est très importante, car il est primordial d’avoir des données cohérentes et fiables.
  Il faut bien sûr s’assurer de la sécurité et de la confidentialité des données que vous utilisez. Plusieurs mesures peuvent être mises en place pour cela :
    - Physique : accès restreint, caméra de sécurité, etc
    - Technologique : cryptographie, mots de passe, firewall etc
      …

  Pour cela, la blockchain est une bonne façon de s’assurer de l’intégrité des données grâce à ses caractéristiques.

- **Pouvoir mettre à disposition les données récoltées à la demande des personnes concernées**

C. Pattern pour contribuer à la privacy

Nous allons ici proposer des patterns pertinents pour contribuer à la privacy des données.

L'utilisation de pattern est intéressante à plusieurs niveaux. Les patterns sont une solution à un problème récurrent dans la conception et la modélisation d'architecture logiciel.

Ils regroupent un ensemble de connaissances et d'expertises acquises par les développeurs et permettent d'exploiter cette connaissance pour la conception de notre solution.

Ils ont pour avantages de répondre à différents types de problème avec une solution approuvée et utilisée par des experts. Ils peuvent également permettre de prévenir de futurs problèmes avant même d'y avoir pensé au préalable. Cela permet de gagner en rapidité, en efficacité et en qualité de conception.

Dès que nous sommes confrontés à un flux de données sortant, il peut y avoir des problèmes de privacy par exemple causés par une dissémination de données privées.
Nous pouvons par ailleurs avoir quelques problèmes quand on parle de stockage de données lié au devoir de respect du droit à l'oubli.

Les patterns proposés seront définis ci-dessous à l'aide de la revue systématique de littérature proposée par Nicolas Six & Nicolas Herbaut & Camille Salinesi [64] [65].

- **Embedded Permission :** Les contrats intelligents utilisent un contrôle de permission intégré pour restreindre l'accès à l'invocation des fonctions définies dans les contrats intelligents. Cela sert à limiter l'appel de fonctions individuelles dans le contrat intelligent à un ensemble de comptes autorisés en intégrant des contrôles d'autorisation.

  Ce pattern peut s'avérer très utile pour la gestion de l'accès aux données et répondre à l'exigence consistant à limiter l'accès aux données de santé des patients.

- **Off-chain data storage** : Pour rappel, ce pattern consiste à stocker les données en dehors de la blockchain. Il permet de réduire les coûts de stockage, mais en préservant l'intégrité des données.

  Ce pattern peut s'avérer très utile pour garantir l'immuabilité et l'intégrité des données pour les grandes données.

- **Encryption on-chain data :** Permet d'assurer la confidentialité des données stockées sur la blockchain en les cryptant. En utilisant le cryptage, les informations accessibles au public sur la blockchain sont cryptées, empêchant quiconque sans la clé secrète d'interpréter les informations.

Il est généralement utilisé en accouplement avec le pattern Off-chain data storage car ils se complètent très bien pour stocker les données hors chaîne tout en préservant la confidentialité des données.

Ce pattern peut s'avérer très utile pour préserver la confidentialité des participants impliqués, pour crypter les données avant de les insérer dans la blockchain et répondre à l'exigence consistant à respecter les règles de sécurité pour protéger les données des patients contre tout accès non autorisé.

- **Time-Constrained Access** : Utilisé quand l'accès au contenu doit être limité dans la fenêtre de temps spécifiée. Il s'utilise nécessairement avec le pattern **Reverse Oracle** pour fournir et limiter l'accès aux informations d'identification stockées sur la chaîne.

  Après modification pour adhérer de meilleures façons au cas d'utilisation liée à la privacy. Ce pattern peut s'avérer utile pour résoudre le problème que pose la blockchain avec le respect du droit à l'oubli.

- **Reverse Oracle** : Utilisé conjointement avec le pattern **Time-Constrained Access.** Avec Reverse oracle, les composants hors chaîne peuvent interroger la blockchain. Les composants hors chaîne d'un système existant reposent sur des contrats intelligents exécutés sur une blockchain pour vérifier les conditions requises et fournir les données demandées.

Bien sûr, il n'existe pas de pattern qui puisse couvrir tous les cas de confidentialité. Dans la pratique, il peut s'avérer intéressant de faire un DPIA (data protection impact assessment) et sur la base de vos exigences, définir quel type de pattern, vous pouvez utiliser.

V. Discussion

Quand on commence à parler de privacy, cela présente un impact important et de grande ampleur dans la conception d'une solution et de l'architecture d'une application.

Auparavant, ne disposant pas de restrictions mises en place, les entreprises disposaient des données personnelles à des finalités non explicites, sans consentement ou encore en les sous-louant à différentes entreprises.

De plus en plus de lois et réglementations sont mises en place pour légiférer sur l'utilisation et la disposition de données confidentielles.

Quand on parle de données médicales, on parle de données très sensibles et très confidentiels. Elles sont donc soumises aux différentes contraintes et réglementations liées aux données personnelles.

Quand la blockchain est utilisée pour traiter des données de santé, elle est tout autant confrontée aux restrictions liées aux données personnelles et doit de ce fait se mettre en conformité avec les réglementations en vigueur.

Grâce à ses caractéristiques, elle a de nombreux avantages pour préserver la sécurité des données. Mais la confidentialité des données est l'une des principales difficultés de la blockchain, également causée par ses caractéristiques. Toutes les informations sur une blockchain sont disponibles pour les participants du réseau.

De plus, la blockchain étant un réseau décentralisé, il n'y a pas d'utilisateur privilégié pouvant par exemple réguler certaines choses dans ce réseau, peu importe que la blockchain soit publique, en consortium ou privée.

C'est pour quoi, nous allons justement dans cette section essayer de répondre la problématique centrale de notre SLR :

Comment est abordée la notion de confidentialité dans l'échange de données médicales utilisant la blockchain ?

Pour répondre à cette question de recherche, nous avons abordé et analysé la littérature selon différents points. Afin de savoir si les chercheurs s'intéressent à la privacy, nous avons regardé si celle-ci était définie au sein des articles. Ensuite, nous avons cherché à connaître le niveau de preuve fourni par les auteurs pour démontrer leurs contributions à la privacy. Puis, nous nous sommes intéressés à la présence de notion légale dans les articles sélectionnés. Enfin, nous avons vérifié si la solution était en conformité avec des réglementations existantes.

Le niveau de définition de la privacy dans la littérature proposant des solutions utilisant la blockchain pour l'échange de données médicales.

Nous nous sommes intéressés à cet aspect, car elle nous permet de visualiser si la privacy est prise en considération dans l'article. Avec seulement quatre articles sur trente-huit, nous avons constaté que la privacy est très peu définie au sein des articles. Ceci nous laisse penser que l'intérêt des auteurs ne s'est pas suffisamment axé sur celle-ci et que toutes les contraintes qu'elle implique n'ont pas été réfléchies pendant la conception de la solution.

Le niveau de preuve fourni par la littérature est insuffisant.

Tout comme la privacy n'est pas suffisamment définie au sein des publications, le niveau de preuve fourni est souvent peu, si ce n'est nullement présent. Les auteurs se contentent de mentionner le fait que les données sont cryptées pour assurer le niveau de privacy.
Comme on l'a compris en définissant la privacy dans la section 2, cela n'est pas suffisant prouver la conformité d'une solution à la privacy.

Les publications font rarement références à des notions légales

Pour continuer dans le fait que le niveau de preuve fourni pour appuyer sa conformité à la privacy au sein des publications est insuffisant, nous constatons que très peu d'articles ne mentionnent des lois ou réglementation liée à la privacy. En effet, on dénombre seulement cinq articles sur trente-huit. Cependant, nous constatons que les articles mentionnant des notions légales démontrent explicitement que leur solution est en conformité avec la privacy. Il semble donc intéressant que les auteurs cherchent à faire références à des notions légales dans leurs articles.

S'occuper de la privacy dès la conception et la modélisation de la solution

Il peut être compliqué de visualiser et analyser la privacy dans la modélisation d'une solution, car elle requiert de nombreuses connaissances. Comme nous l'avons vu, l'utilisation de Linddun Go permet de pallier cette difficulté en se posant les bonnes questions sur les différentes menaces potentiellement problématiques.
Cela nous permet alors d'analyser la privacy dès la conception de la solution, et ce, sans connaissance poussée de celle-ci.

Pour autant, les solutions respectent en parties les exigences liées à la privacy grâce à la blockchain

Nous avons constaté que le fait que la privacy ne soit pas suffisamment prise en considération ne signifie pas forcément que la solution ne respecte pas les critères énoncés par les réglementations. En effet, avec trente-trois publications sur trente-huit, la majorité des articles respectent des exigences liées à des lois ou réglementations existantes. Cela est lié à l'utilisation de la blockchain qui, grâce à ses caractéristiques, fait en sorte que la solution soit conforme à quelques critères liés à la privacy.

La blockchain, une solution prometteuse pour les DSE

Comme nous venons de le voir, la blockchain permet de se conformer à certaines exigences soumises par la privacy. Voici quelques exemples des règles concernant la protection des données avec la RGPD :

- Respecter des règles de sécurité pour protéger les données des patients contre des accès non autorisés
- Tenir un registre des activités
- Prioriser l'utilisation de blockchain à permission :
    - Limiter l'accès aux données de santé des patients : seules certaines personnes sont autorisées, au regard de leurs missions, à accéder à celles-ci
- Assurez l'intégrité des données

Toutes les publications extraites dans le cadre de notre revue systématique de la littérature utilisent une blockchain de type privée ou à permission, car une blockchain publique est par défaut totalement incompatible avec la privacy et le partage de données de santé.

Ces critères liés à la privacy sont par défaut respecté en utilisant la blockchain. La blockchain sécurise les données en chiffrant les données via le hachage et la cryptographie. Elle permet de tenir un registre par défaut et l'utilisation de blockchain à permission permet de limiter l'accès à la blockchain.

Cependant, un problème existe concernant le droit à l'oubli imposant le fait que les données doivent pouvoir être supprimées et effacées au bout d'un certain temps. Ce problème est analysé au sein d'une publication et les chercheurs ont fourni une solution pour ce problème. Des solutions sont donc envisageables pour se conformer au droit à l'oubli [12].

VI. Conclusion

La blockchain est une technologique récente qui embarque de nombreuses caractéristiques intéressantes, notamment dans le secteur de la santé.

Dans un environnement où nous devons garantir la confidentialité, la sécurité, l'interopérabilité, la traçabilité et l'intégrité des données sensibles que représentent les données de santé, les solutions misent en place doivent garantir ces différentes exigences.

La blockchain semble, sur le papier, être une bonne candidate pour le respect de ces contraintes. La blockchain fournie un support compatible avec toutes les caractéristiques citées ci-dessus : intégrité, traçabilité, sécurité…

Mais la blockchain comprend également d'autre attribut tel que l'immutabilité et la transparence des données. Toutes les données stockées sur la blockchain sont visibles par tous les participants de celle-ci et ces données ne sont ni modifiables si supprimables.

Ces deux points sont en contradiction avec les exigences que requiert la privacy pour les données personnelles. Des solutions pour pallier ces incompatibilités peuvent être développées.

Cependant, il s'agit de savoir si les auteurs proposant des solutions pour l'échange de donnée médicale avec la blockchain ont connaissances de ces différentes problématiques et de l'ampleur des exigences qu'impliquent la privacy et la protection des données personnelles.

Ainsi, notre revue littéraire systématique a pour but d'étudier cette problématique afin en analysant les publications dans la littérature existante. Pour cela, on s'intéresse à savoir si les publications décrivent la privacy ainsi que le niveau de preuve qu'elles fournissent. Cela nous sert à recenser le niveau d'intérêt que porte les auteurs sur la privacy et donc leur effort pour se mettre en conformité avec celle-ci.

On remarque dans notre base de publications que 89% des publications ne définissent pas la privacy et que 87% ne mentionnent pas de lois, juridiction ou réglementation existante qui encadre le respect des données personnelles. De plus, on constate que les articles dans cette situation s'intéressent à la sécurité des informations, mais pas forcément à la privacy de celle-ci.

Même si ces deux aspects sont complémentaires, il est important de faire la différence, car leurs exigences peuvent se rejoindre, mais ne sont pas identiques :

Voici la différence entre ces deux notions selon la CNIL :
"L'objectif de la sécurité de l'information est de protéger l'organisme des atteintes liées à son patrimoine informationnel. Celui de la protection de la vie privée est de protéger les personnes des atteintes liées à leurs données." [72].

Les chercheurs doivent donc prendre plus en considération l'aspect de la privacy dans le cahier des charges de leurs applications.

Ainsi, il serait intéressant d'instaurer l'utilisation d'outils de modélisation de menaces à la vie privée tels que Linddun pour analyser, avec des experts en privacy, les failles d'une application sous une vision "privacy compliance".

Pour finir, ce travail fourni quelques points visant la conformité à la privacy dont une checklist des choses à prendre en compte quand on traite la privacy ainsi que les patterns intéressant pour contribuer à la privacy.

VII. Références

[1] Yang, G., Li, C., & Marstein, K. E. (2019). A blockchain‐based architecture for securing electronic health record systems. Concurrency and Computation: Practice and Experience, 33(14). Portico.

[2] Al Mamun, A., Faruk Jahangir, Md. U., Azam, S., Kaiser, M. S., & Karim, A. (2020). A Combined Framework of InterPlanetary File System and Blockchain to Securely Manage Electronic Medical Records. Proceedings of International Conference on Trends in Computational and Cognitive Engineering, 501–511.

[3] Li, J. (2020). A New Blockchain-based Electronic Medical Record Transferring System with Privacy. 2020 5th International Conference on Information Science, Computer Technology and Transportation (ISCTT).

[4] Centobelli, P., Cerchione, R., & Riccio, E. (2021). A novel architecture for enhancing Electronic Health Record interoperability: A Blockchain-based approach. 2021 IEEE Technology & Engineering Management Conference - Europe (TEMSCON-EUR).

[5] Rahman, M. S., Khalil, I., Mahawaga Arachchige, P. C., Bouras, A., & Yi, X. (2019). A Novel Architecture for Tamper Proof Electronic Health Record Management System using Blockchain Wrapper. Proceedings of the 2019 ACM International Symposium on Blockchain and Secure Critical Infrastructure- BSCI ’19.

[6] Ali, A., Rahim, H. A., Ali, J., Pasha, M. F., Masud, M., Rehman, A. U., Chen, C., & Baz, M. (2021). A Novel Secure Blockchain Framework for Accessing Electronic Health Records Using Multiple Certificate Authority. Applied Sciences, 11(21), 9999.

[7] Adlam, R., & Haskins, B. (2019). A Permissioned Blockchain Approach to the Authorization Process in Electronic Health Records. 2019 International Multidisciplinary Information Technology and Engineering Conference (IMITEC).

[8] Guo, H., Li, W., Nejad, M., & Shen, C.-C. (2019). Access Control for Electronic Health Records with Hybrid Blockchain-Edge Architecture. 2019 IEEE International Conference on Blockchain (Blockchain).

[9] Kiran Dash, K., Nayak, B., & Kumar Mohanta, B. (2021). An Approach to Securely Store Electronic Health Record(EHR) Using Blockchain with Proxy Re-Encryption and Behavioral Analysis. Advances in Intelligent Systems and Computing, 415–423.

[10] Tang, F., Ma, S., Xiang, Y., & Lin, C. (2019). An Efficient Authentication Scheme for Blockchain-Based Electronic Health Records. IEEE Access, 7, 41678–41689.

[11] Verdonck M, Poels G. Architecture and Value Analysis of a Blockchain-Based Electronic Health Record Permission Management System. In: Proceedings of 14th International Workshop on Value Modelling and Business Ontologies.: Ghent University; 2020 Presented at: 14th International Workshop on Value Modelling and Business Ontologies; Jan 16, 2020; Brussels, Belgium URL: https://biblio.ugent.be/publication/8654183

[12] Xia, Q., Sifah, E., Smahi, A., Amofa, S., & Zhang, X. (2017). BBDS: Blockchain-Based Data Sharing for Electronic Medical Records in Cloud Environments. Information, 8(2), 44.

[13] Ray, P. P., Chowhan, B., Kumar, N., & Almogren, A. (2021). BIoTHR: Electronic Health Record Servicing Scheme in IoT-Blockchain Ecosystem. IEEE Internet of Things Journal, 8(13), 10857–10872.

[14] Ajayi, O., Abouali, M., & Saadawi, T. (2020). Blockchain Architecture for Secured Inter-healthcare Electronic Health Records Exchange. Advances in Intelligent Systems and Computing, 161–172.

[15] Alexaki, S., Alexandris, G., Katos, V., & Petroulakis, N. E. (2018). Blockchain-based Electronic Patient Records for Regulated Circular Healthcare Jurisdictions. 2018 IEEE 23rd International Workshop on Computer Aided Modeling and Design of Communication Links and Networks (CAMAD).

[16] Liu, J., Li, X., Ye, L., Zhang, H., Du, X., & Guizani, M. (2018). BPDS: A Blockchain Based Privacy-Preserving Data Sharing for Electronic Medical Records. 2018 IEEE Global Communications Conference (GLOBECOM).

[17] Kalaipriya, R., Devadharshini, S., Rajmohan, R., Pavithra, M., & Ananthkumar, T. (2020). Certain Investigations on Leveraging Blockchain Technology for Developing Electronic Health Records. 2020 International Conference on System, Computation, Automation and Networking (ICSCAN).

[18] Wang, Y., & He, M. (2021). CPDS: A Cross-Blockchain Based Privacy-Preserving Data Sharing for Electronic Health Records. 2021 IEEE 6th International Conference on Cloud Computing and Big Data Analytics (ICCCBDA).

[19] Zhang, J., Li, Z., Tan, R., & Liu, C. (2021). Design and Application of Electronic Rehabilitation Medical Record (ERMR) Sharing Scheme Based on Blockchain Technology. BioMed Research International, 2021, 1–12.

[20] Sari, P. K., & Yazid, S. (2020). Design of Blockchain-based Electronic Health Records for Indonesian Context: Narrative Review. 2020 International Workshop on Big Data and Information Security (IWBIS).

[21] Kim, M., Yu, S., Lee, J., Park, Y., & Park, Y. (2020). Design of Secure Protocol for Cloud-Assisted Electronic Health Record System Using Blockchain. Sensors, 20(10), 2913.

[22] Riadi, I., Ahmad, T., Sarno, R., Purwono, P., & Ma'arif, A. (2022). Developing Data Integrity in an Electronic Health Record System using Blockchain and InterPlanetary File System (Case Study: COVID-19 Data). Emerging Science Journal, 4, 190–206.

[23] Niu, S., Li, W., & Liu, W. (2020). Electronic Health Record Data Sharing Cryptographic Algorithm Based on Blockchain. Artificial Intelligence and Security, 363–375.

[24] Niu, S., Chen, L., Wang, J., & Yu, F. (2020). Electronic Health Record Sharing Scheme With Searchable Attribute-Based Encryption on Blockchain. IEEE Access, 8, 7195–7204.

[25] Wu, S., & Du, J. (2019). Electronic medical record security sharing model based on blockchain. Proceedings of the 3rd International Conference on Cryptography, Security and Privacy - ICCSP '19.

[26] Martinez, A., Molina, C., & Subauste, D. (2020). Electronic Medical Records Management in Health Organizations using a Technology Architecture based on Blockchain. 2020 IEEE ANDESCON.

[27] Alfaidi, A., & Chow, E. (2020). Health Record Chain (HRC): Implementation of Mobile Healthcare system using Blockchain to enhance Privacy of Electronic Health Record EHR. 2020 International Conference on Computational Science and Computational Intelligence (CSCI).

[28] Gutiérrez, O., Romero, G., Pérez, L., Salazar, A., Charris, M., & Wightman, P. (2020). HealthyBlock: Blockchain-Based IT Architecture for Electronic Medical Records Resilient to Connectivity Failures. International Journal of Environmental Research and Public Health, 17(19), 7132.

[29] Arul, P., & Renuka, S. (2021). Hyperledger blockchain based secure storage of electronic health record system in edge nodes. Journal of Physics: Conference Series, 2115(1), 012034.

[30] Uddin, M., S. Memon, M., Memon, I., Ali, I., Memon, J., Abdelhaq, M., & Alsaqour, R. (2021). Hyperledger Fabric Blockchain: Secure and Efficient Solution for Electronic Health Records. Computers, Materials & Continua, 68(2), 2377–2397.

[31] Hashim, F., Shuaib, K., & Sallabi, F. (2021). MedShard: Electronic Health Record Sharing Using Blockchain Sharding. Sustainability, 13(11), 5889.

[32] Zhao, Y., Cui, M., Zheng, L., Zhang, R., Meng, L., Gao, D., & Zhang, Y. (2019). Research on electronic medical record access control based on blockchain. International Journal of Distributed Sensor Networks, 15(11), 155014771988933.

[33] Mahore, V., Aggarwal, P., Andola, N., Raghav, & Venkatesan, S. (2019). Secure and Privacy Focused Electronic Health Record Management System using permissioned Blockchain. 2019 IEEE Conference on Information and Communication Technology.

[34] Guo, R., Shi, H., Zhao, Q., & Zheng, D. (2018). Secure Attribute-Based Signature Scheme With Multiple Authorities for Blockchain in Electronic Health Records Systems. IEEE Access, 6, 11676–11686.

[35] Beinke, J. H., Fitte, C., & Teuteberg, F. (2019). Towards a Stakeholder-Oriented Blockchain-Based Architecture for Electronic Health Records: Design Science Research Study. Journal of Medical Internet Research, 21(10), e13585.

[36] Chenthara, S., Ahmed, K., Wang, H., Whittaker, F., & Chen, Z. (2020). Healthchain: A novel framework on privacy preservation of electronic health records using blockchain technology. PLOS ONE, 15(12), e0243043. https://doi.org/10.1371/journal.pone.0243043

[37] Pilares, I. C. A., Azam, S., Akbulut, S., Jonkman, M., & Shanmugam, B. (2022). Addressing the Challenges of Electronic Health Records Using Blockchain and IPFS. Sensors, 22(11), 4032. https://doi.org/10.3390/s22114032

[38] Mamun, A. A., Azam, S., & Gritti, C. (2022). Blockchain-Based Electronic Health Records Management: A Comprehensive Review and Future Research Direction. IEEE Access, 10, 5768–5789. https://doi.org/10.1109/access.2022.3141079

[46] Sookhak, M., Jabbarpour, M. R., Safa, N. S., & Yu, F. R. (2021). Blockchain and smart contract for access control in healthcare: A survey, issues and challenges, and open issues. Journal of Network and Computer Applications, 178, 102950. https://doi.org/10.1016/j.jnca.2020.102950

[52] Barbara Kitchenham & Stuart Charters & David Budgen &r Pearl Brereton & Mark Turner & Steve Linkman & Magne Jørgensen & Emilia Mendes & Giuseppe Visaggio (2007)

[58] Wieringa, R., Maiden, N., Mead, N., & Rolland, C. (2005). Requirements engineering paper classification and evaluation criteria: a proposal and a discussion. Requirements Engineering, 11(1), 102–107. https://doi.org/10.1007/s00766-005-0021-6

[59] Uzunov, A. V., & Fernandez, E. B. (2014). An extensible pattern-based library and taxonomy of security threats for distributed systems. Computer Standards & Interfaces, 36(4), 734–747. https://doi.org/10.1016/j.csi.2013.12.008

[60] Bedi, P., Gandotra, V., Singhal, A., Narang, H., & Sharma, S. (2012). Threat-oriented security framework in risk management using multiagent system. Software: Practice and Experience, 43(9), 1013–1038. https://doi.org/10.1002/spe.2133

[61] Baquero, A.O., Kornecki, A.J., Zalewski, J.: Threat modeling for aviation computer security. Fusing IT Real-Time Tactical. 28, 21–27 (2015)

[64] Six, N., Herbaut, N., & Salinesi, C. (2022). Blockchain software patterns for the design of decentralized applications: A systematic literature review. Blockchain: Research and Applications, 3(2), 100061. https://doi.org/10.1016/j.bcra.2022.100061

VIII. Webographie

[39] https://www.ipsos.com/fr-fr/sante-lappel-des-francais-aux-candidats-la-presidentielle

[40] https://www.cnil.fr/fr/rgpd-de-quoi-parle-t-on

[41] https://oag.ca.gov/privacy/ccpa

[42] https://lgpd-brazil.info/

[43] https://usercentrics.com/knowledge-hub/china-personal-information-protection-law/

[44] https://www.proofpoint.com/fr/threat-reference/hipaa-compliance

[45] https://esante.gouv.fr/labels-certifications/hds/certification-des-hebergeurs-de-donnees-de-sante

[47] https://www.cnil.fr/fr/rgpd-et-professionnels-de-sante-liberaux-ce-que-vous-devez-savoir

[48] https://www.cnil.fr/fr/lacces-au-dossier-medical

[49] https://www.cnil.fr/fr/definition/responsable-de-traitement#:~:text=Le%20responsable%20de%20traitement%20est,incarn%C3%A9e%20par%20son%20repr%C3%A9sentant%20l%C3%A9gal.

[50] https://www.cnil.fr/sites/default/files/atoms/files/la_blockchain.pdf

[51] https://www.healthit.gov/faq/what-electronic-health-record-ehr

[53] https://parsif.al/about/

[54] https://www.cnil.fr/fr/blockchain-et-rgpd-quelles-solutions-pour-un-usage-responsable-en-presence-de-donnees-personnelles

[55] https://www.cnil.fr/fr/reglement-europeen-protection-donnees/chapitre3#Article17

[56] https://www.cointribune.com/guides-crypto/bien-debuter/quest-ce-que-le-crosschain/

[57] https://www.xenon-360.fr/differents-types-de-blockchain/

[62] https://www.linddun.org/linddun

[63] https://distrinet.cs.kuleuven.be/software/linddungo/

[65] https://beta.ontotool.nicosix.com/

[66] https://www.cnil.fr/fr/travailler-avec-un-sous-traitant-dans-une-collectivite

[67] https://www.cnil.fr/fr/RGPD-analyse-impact-protection-des-donnees-aipd

[68] https://data-droit.com/2019/06/12/que-signifient-les-principes-de-privacy-by-design-et-privacy-by-default/

[69] https://ipfs.tech/

[70] https://www.investopedia.com/terms/h/hash.asp

[71] https://sion.info/assets/pdf/publications/WuytsIWPE2020.pdf

[72] https://www.cnil.fr/fr/garantir-la-securite-des-donnees#:~:text=Les%20deux%20logiques%20sont%20compl%C3%A9mentaires,atteintes%20li%C3%A9es%20%C3%A0%20leurs%20donn%C3%A9es.

[73] https://www.linddun.org/_files/ugd/cc602e_cf7e4c6b1d894bdaabc3094c48b26869.pdf

IX. Tables des annexes

X. Tables des figures

XI. Tables des tableaux

XII. Glossaire

Responsables de traitement : “Le responsable de traitement est la personne morale (entreprise, commune, etc.) ou physique qui détermine les finalités et les moyens d’un traitement, c’est à dire l’objectif et la façon de le réaliser. En pratique et en général, il s’agit de la personne morale incarnée par son représentant légal.” [49]

Sous-traitant des données : Le sous-traitant, au sens du RGPD, est la personne physique ou morale (entreprise ou organisme public) qui traite des données pour le compte d’un autre organisme (le responsable de traitement), dans le cadre d’un service ou d’une prestation. [66]

DPIA (Data Protection Impact Assessment) : L’AIPD est un outil qui permet de construire un traitement conforme au RGPD et respectueux de la vie privée. Elle concerne les traitements de données personnelles qui sont susceptibles d'engendrer un risque élevé pour les droits et libertés des personnes concernées. [67]

Privacy by design : Ce concept prévoit que les considérations de protection de la vie privée doivent être prises en compte dès la conception du produit ou du service amené à collecter, traiter ou utiliser des données personnelles. C'est-à dire que le respect de la vie privée doit dès le départ constituer une préoccupation des développeurs afin de l’intégrer dans la structure même du service. [68]

Privacy by default : Ce concept implique que dès la conception du service ou du produit, le plus haut niveau de protection de la vie privée soit mis en place et applicable par défaut. C’est-à-dire que sans aucune intervention de la part de l’utilisateur, l’ensemble des mesures disponibles afin de protéger les données personnelles et d’en limiter la collecte doit être activé. [68]

IPFS (InterPlanetary File System) : Un protocole hypermédia peer-to-peer conçu pour préserver et développer les connaissances de l'humanité en rendant le Web évolutif, résilient et plus ouvert. [69]

Hash (fonction de hashage) : Un hachage est une fonction mathématique qui convertit une entrée de longueur arbitraire en une sortie chiffrée de longueur fixe. Ainsi, quelle que soit la quantité initiale de données ou la taille du fichier impliqué, son hachage unique sera toujours de la même taille. [70]

XIII. Déclaration d'intérêts concurrents

Les auteurs déclarent ne pas avoir d'intérêts financiers concurrents ou de relations personnelles connus qui auraient pu sembler influencer les travaux rapportés dans cet article.