Le chaînage de bases de données anonymisées pour les études épidémiologiques multicentriques nationales et internationales : proposition d'un algorithme cryptographique

Catherine QUANTIN, Maniane FASSA[1] et Gilles TROUESSIN[2]

Objectif : Pour conduire des études épidémiologiques multicentriques nationales ou internationales, il est souvent nécessaire de rapprocher des informations d'un même patient, provenant de plusieurs sources. En Europe, le chaînage des fichiers nominatifs, dans le cadre de la recherche médicale, est soumis à la directive européenne du 24 octobre 1995, qui requiert que l'information soit rendue anonyme avant son utilisation à des fins de chaînage. La méthodologie du hachage permet de résoudre le problème de l'anonymisation des données, notamment en santé. Par ailleurs, pour des raisons de sécurité, il est recommandé d'utiliser des clés différentes pour chaque étude. Malheureusement, cette recommandation est en contradiction avec les besoins de chaînage. L'objectif de cet article est de proposer une méthodologie innovante pour répondre à la fois aux exigences en matière de sécurité des informations médicales, tout en permettant le chaînage des données relatives à un même patient et leur exploitation statistique.

Méthodes : La méthode repose sur l'utilisation, pour le hachage, de la fonction SHA (Secure Hash Algorithm), qui permet d'assurer l'anonymat des données personnelles, qui est combinée avec des techniques de chiffrement. L'originalité de la méthode réside dans la manière dont le hachage et le chiffrement sont combinés : comme dans les méthodes de chiffrement asymétrique, nous proposons l'utilisation de deux clés, mais avec une différence fondamentale puisqu'une des deux clés va dépendre de l'identité du patient.

Résultats : La combinaison du hachage et des techniques cryptographiques assure une amélioration importante dans la sécurité des données, tout en permettant le chaînage des données multicentriques.

Conclusion : Cette méthode rend disponibles les informations rendues anonymes et stockées dans des bases de données multicentriques nationales et internationales, pour une exploitation à des fins épidémiologiques et de recherche clinique. Ceci, en respectant les exigences de sécurité imposées par les lois nationales et européennes.

[1] Inserm et CHU de Dijon

[2] Oppida-Sud, Toulouse