MÉTHODOLOGIE POUR LE CHAINAGE DE DONNÉES SENSIBLES TOUT EN RESPECTANT L'ANONYMAT : APPLICATION AU SUIVI DES INFORMATIONS MÉDICALES

C. QUANTIN[1]
En collaboration avec : F-A. ALLAERT [1,2], *R. PATTISINA* [1], *C. BINQUET*[1], *B. CORNET*[3], *J-B. GOUYON*[3], *J-M. RODRIGUES*[4], *L. DUSSERE*[1].

(1) Service de Biostatistique et Informatique Médicale
(2) Chairman TC/251/WGIII Centre Européen de Normalisation, CEN BIOTECH,
(3) Service de Pédiatrie 2, Hôpital d'Enfants
(4) Département de Santé Publique et d'Information, Université de Saint Etienne,

1. Introduction

En dehors des utilisations imposées par l'assurance maladie ou les services de l'état (feuille de soins électroniques, PMSI...) il est possible d'envisager des utilisations et des circulations d'information propres aux médecins par exemple dans le cadre de réseaux de soins. Toutefois, le regroupement des informations médicales relatives à un même patient par le croisement de divers fichiers existants, ne peut aller à l'encontre de la législation européenne et française relative à la protection des libertés individuelles vis à vis du traitement automatisé des données personnelles.

Dans la présentation, nous montrerons que le respect de la législation conduit au paradoxe suivant : pouvoir réunir les différentes parties du dossier d'un même patient sans pouvoir accéder à son identité. Nous verrons comment les techniques cryptographiques, telle que la procédure d'anonymat et de chaînage développée par le Département d'Information Médicale (DIM) du CHU de Dijon, apportent une solution à ce paradoxe.

2. La législation sur la sécurité des informations nominatives

L'article 4 de la loi du 6 janvier 1978 relative à l'informatique, aux fichiers et aux libertés donne la définition des informations nominatives bénéficiant d'une protection au regard des libertés publiques : *« Sont réputées nominatives au sens de la présente loi les informations qui permettent sous quelque forme que ce soit, directement ou non, l'identification des personnes physiques auxquelles elles s'appliquent, que le traitement soit effectué par une personne physique ou par une personne morale. ».* Cette définition introduit ainsi la notion d'informations « indirectement nominatives » qui met sur un pied d'égalité vis à vis de la loi les informations accompagnées « directement » du nom de l'individu auquel elles se rapportent et celles qui ne sont pas accompagnées du nom mais qui font partie d'un faisceau de données convergentes tel qu'il est possible d'identifier l'individu en cause.

Une notion similaire est reprise dans la Directive européenne du 24 octobre 1995 relative à la protection des personnes physiques à l'égard du traitement des données à caractères personnel et à la libre circulation de ces données. Cette directive définit une exigence de protection commune à l'ensemble des pays européens pour contribuer au libre échange des informations sur le marché intérieur de l'Europe et prévoit des dispositions particulières pour le transfert des données vers les pays ne disposant pas de garanties équivalentes à ses propres exigences. Cette directive substitue au concept d'informations nominatives celui de « données *à caractère personnel » : « toute information concernant une personne physique identifiée ou identifiable (personne concernée) ; est réputée identifiable une personne qui peut être identifiée, directement ou indirectement, notamment par référence à un numéro d'identification ou à un ou plusieurs éléments spécifiques propres à son identité physique, physiologique, psychique, économique, culturelle ou sociale.* » Cette définition très exhaustive permet à terme de considérer que toute base de données est indirectement nominative [1].

Cette directive européenne devrait avoir été transcrite en droit français depuis le 24 octobre 1998 et fait l'objet actuellement d'un projet de loi non encore déposé devant le Parlement. De l'ensemble de ces considérations, il résulte que la notion de données nominatives ou personnelles concerne un très grand nombre d'informations même si le nom n'apparaît pas et qu'aucune table de correspondance entre l'identité en clair et des codes alphanumériques substitutifs n'existe. Sur le plan statistique, le risque d'identification d'un individu à partir d'informations apparemment anonymes est loin d'être nul, en raison de la possibilité de croisement avec une grande diversité de fichiers existants ou à venir. Qui aurait pensé, il y a encore peu de temps, que l'identifiant de la Sécurité sociale pourrait être traité informatiquement par les services fiscaux [2] ? Toutefois, ce souci que les personnes concernées ne puissent être identifiées lors de l'évaluation ou lors de l'analyse des pratiques et des

activités de soins et de prévention, est repris intégralement dans l'article 41 de la loi n°99-641 du 27 juillet 1999 portant création d'une couverture maladie universelle modifiant l'article 40-10 de la loi du 6 janvier 1978.

3. Anonymat : l'essor des méthodes cryptographiques

Plutôt que d'utiliser des méthodes statistiques d'anonymat reposant sur la perturbation de données pour empêcher l'identification des personnes et par conséquent induisant une perte de qualité des informations, il paraît préférable d'utiliser des techniques cryptographiques pour assurer la sécurité des informations. Ces méthodes ne sont pas beaucoup plus récentes que les méthodes statistiques d'anonymat mais leur utilisation était restreinte par la loi pour des raisons de défense nationale. Les autorisations d'utilisation de ces méthodes n'étaient donc pas faciles à obtenir auprès du Service Central de la Sécurité des Systèmes d'Informations (SCSSI), service qui dépend directement du 1er Ministre.

Le domaine de la cryptographie a bénéficié assez récemment d'une libéralisation, d'abord en 1998 [3-5] sous forme d'une simplification de la procédure de déclaration auprès du SCSSI. Cet assouplissement en faveur de l'utilisateur, a priori non spécialiste du domaine, a été poursuivi en 1999 [6], en faisant porter le poids de la réglementation sur les professionnels de la cryptologie. En particulier, l'utilisation des clés de haute sécurité d'une longueur de 128 bits a été rendue possible (jusqu'alors limitée à 40 bits) ; cette évolution étant devenue obligatoire pour satisfaire aux exigences de la reconnaissance de la signature électronique et faciliter les transactions commerciales dans le cadre de l'Internet.

Cette libéralisation a permis de lever l'obstacle de l'utilisation des techniques de cryptage pour assurer la confidentialité des informations médicales directement ou indirectement nominatives et appelées à circuler sur des réseaux informatiques. En effet, si la Commission Nationale de l'Informatique et des Libertés (CNIL) accepte des clés de seulement 40 bits pour le cryptage des informations indirectement nominatives, elle exige des clés d'une longueur d'au moins 56 bits pour celles qui sont directement nominatives.

Si l'on s'intéresse à la sécurité des informations médicales dans le cadre d'un réseau, les méthodes de cryptage peuvent être utilisées à trois niveaux (Figure 1). Le premier niveau concerne le respect de la confidentialité des informations pendant leur transmission. La confidentialité [7], selon la définition donnée par le Centre Européen de Normalisation[1], est assurée lorsque seuls les utilisateurs dûment habilités ont accès à l'information. Il s'agit par exemple d'un échange d'information

[1] Groupe de travail « Qualité et Sécurité » Working Group III « Quality and Security » de la Communauté Européenne.

entre un établissement de santé et un cabinet médical libéral. Chiffrer un message, c'est lui appliquer une fonction de transformation qui le rendra illisible au tout-venant. Cette fonction est appliquée par un algorithme de chiffrement ou de cryptage. En règle générale, l'algorithme est public, et la confidentialité n'est assurée que par la clé de l'utilisateur [8-10], qui doit donc être difficile à retrouver, même pour un cryptanalyste expérimenté. Le médecin hospitalier, grâce à l'utilisation d'une méthode de cryptage, sera sûr que seul le médecin généraliste auquel ce message est destiné pourra en prendre connaissance, puisqu'en tant que destinataire légitime il sera le seul à connaître la clé de déchiffrement.

Le deuxième niveau concerne l'utilisation des méthodes de signature numérique pour permettre au médecin receveur d'authentifier le médecin émetteur du message. Dans l'exemple que nous venons de prendre, ceci signifie que le médecin généraliste pourra s'assurer que le message a bien été adressé par le médecin hospitalier annoncé. L'utilisation de la signature numérique va permettre également de garantir l'intégrité du message, c'est à dire être sûr que le message n'a pas été modifié pendant sa transmission. La signature numérique vient d'être reconnue comme ayant valeur légale par la loi française n°2000-230 du 13 mars 2000 portant adaptation du droit de la preuve aux technologies de l'information et relatif à la signature électronique.

Le troisième niveau d'utilisation des techniques cryptographiques concerne le regroupement d'informations médicales au sein d'une structure extérieure aux soins. Le problème du chaînage d'informations médicales nominatives pour la mise en œuvre d'études épidémiologiques multicentriques se pose de plus en plus fréquemment, par exemple dans le cadre d'études coopératives ville/hôpital. Selon les recommandations de la CNIL, il est alors préférable d'utiliser des techniques cryptographiques garantissant une transformation irréversible des données.

Pour résoudre le paradoxe de répondre conjointement aux nécessités de l'anonymat et du rapprochement des informations concernant un même patient, divers outils ont été développés sur le même principe de base :

- le DIM du C.H.U. de DIJON a développé à partir de 1995 une procédure d'ANONYMAT et de chaînage [11,12] beaucoup plus simple et donc plus souple d'utilisation que celle de Michaelis et al [13]. Celle-ci a été déclarée auprès de la CNIL et du Service Central de la Sécurité des Systèmes d'Information (SCSSI) en mars 1996, conformément à la législation (loi n° 90-1170 du 29 décembre 1990 [14] modifiée par la loi n° 96-659 du 26 juillet 1996 [15] sur l'utilisation des procédés cryptographiques). Le procédé cryptographique utilisé fournit une transformation irréversible de l'identité tout en permettant le rapprochement des informations d'un même patient, puisque le code

résultant de cette transformation est toujours le même pour une identité donnée.

- Le CESSI/CNAMTS[2] « a conçu et fourni dès 1996, pour la mise en place du PMSI établissements privés, sur recommandation de la CNIL (qui a suggéré l'utilisation de l'algorithme développé par le DIM du CHU de Dijon), sur demande du MES/DH[3] et après expertise du SCSSI, une fonction d'anonymisation FOIN (Fonction d'Occultation d'Information Nominatives) [16], permettant de remplacer l'identité des patients par des numéros d'anonymat, ou clés de chaînage, pérennes dans le temps (tant que le secret de la fonction à sens unique est inchangé) et dans l'espace (pour chacune des quelques 1000 cliniques privées) » [17].

4. Une procédure pour assurer conjointement l'anonymat et le chaînage des informations médicales

La procédure d'ANONYMAT et de chaînage, développée au DIM du CHU de Dijon, se déroule en deux temps. Une première étape concerne la transformation irréversible des variables d'identification [18] (nom, prénom, date de naissance, sexe, ...) pour obtenir un code strictement anonyme, qui constitue le repère de chaînage. La seconde étape est celle du croisement des fichiers pour chaîner les informations d'une même personne.

4.1 Algorithme de hachage

Contrairement au cryptage [9,19,20] qui doit pouvoir être réversible, notamment lors du déchiffrement du message par son destinataire légitime, les méthodes de « hachage » à sens unique sont mathématiquement irréversibles. Ces algorithmes de compression numérique irréversible sont utilisés notamment pour les signatures électroniques. En accord avec le SCSSI, nous avons choisi le Standard Hash Algorithm (SHA) qui, à notre connaissance, est l'algorithme du domaine public le plus sûr vis à vis des tentatives de déchiffrement [20].

[2] Centre d'Etudes des Sécurités du Système d'Information (CESSI) de la Caisse Nationale de l'Assurance Maladie des Travailleurs Salariés (CNAMTS).

[3] Direction des Hôpitaux (DH) au Ministère de l'Emploi et de la Solidarité (MES), à l'époque Ministère du Travail et des Affaires Sociales (MTAS).

Bien qu'irréversible, l'opération de hachage ne garantit pas la sécurité parfaite des informations. L'algorithme étant public, le hachage pourrait être appliqué à un grand nombre d'identités. On pourrait alors confronter les codes obtenus au code d'un individu donné du fichier haché et retrouver ainsi son identité. On parle alors « d'attaque par dictionnaire ».

Pour prévenir ce type d'attaque, deux clés informatiques sont utilisées (Figure 2). La modification introduite par chaque clé varie d'une identité à l'autre mais est toujours la même pour une identité donnée. La première clé, k1, utilisée au moment du hachage des identités par chaque centre de recueil des informations, permet de protéger les informations vis à vis des personnes qui ne connaissent pas cette clé. Pour assurer la sécurité des informations vis à vis des centres de recueil détenteurs de la clé k1, les informations reçues par le centre de traitement assurant le croisement des fichiers sont à nouveau hachées par le même algorithme mais avec une seconde clé k2. A l'issue des deux hachages des données d'identité, réalisés successivement au niveau des centres de recueil et du centre de traitement, l'anonymat des fichiers est ainsi définitivement préservé.

4.2 Algorithme de chaînage

L'objectif du chaînage est de confronter des fichiers doublement hachés provenant de sources différentes, pour associer les observations qui se rapportent à un même individu. Deux types d'erreurs [21] peuvent survenir dans le processus de chaînage :

Le premier correspond au chaînage de deux observations concernant deux individus différents et constitue une erreur « d'homonymie » : par exemple si l'on associe à tort des informations concernant deux personnes dénommées respectivement Dupond et Dupont, du fait d'une erreur dans la saisie de leurs identités.
Le deuxième type d'erreur correspond à l'absence de chaînage de deux observations d'un même individu et constitue l'erreur « de synonymie » : par exemple en cas d'utilisation successive du nom de jeune fille et du nom marital pour la même femme.

Ces erreurs pourraient être dues soit à des erreurs dans le recueil des données d'identité, soit à la méthode de hachage elle-même. En particulier, des erreurs d'homonymie pourraient résulter de l'existence de collisions lors du hachage, c'est à dire de l'obtention du même code à partir du hachage de deux identités différentes. Dans le cas de l'algorithme SHA retenu pour la procédure de hachage, il s'avère que le taux de collisions est très faible (de l'ordre de 10^{-48}) et que le risque d'erreur d'homonymie correspondant est négligeable [22 p. 97].

Pour réduire l'impact des erreurs de saisie de l'identité sur le chaînage, un traitement orthographique a été intégré dans la procédure d'anonymat. La méthode de chaînage

" AUTOMATCH " proposée par JARO [23] et très utilisée aux USA [24], a été adaptée. Elle tient compte simultanément de plusieurs variables d'identification : le nom, le prénom, le nom de jeune fille, la date de naissance, le sexe et le code postal du lieu de résidence. Bien sûr, chacune de ces variables n'identifie pas un individu de manière pathognomonique, et l'on est ramené au problème connu de la valeur informationnelle d'un signe. Chaque variable est alors pondérée en fonction de la quantité d'information qu'elle apporte. Par exemple, on attribue une valeur plus importante à l'information fournie par la date de naissance qu'à celle fournie par le sexe. Pour déterminer si deux observations doivent être chaînées, on applique un modèle d'analyse statistique qui tient compte des coefficients de pondération de chaque variable utilisée.

5. Premières applications de la procédure conjointe d'anonymat et de chaînage

5.1 Etude de l'interfile active des personnes cancéreuses de trois structures hospitalières dans le cadre de la planification régionale en Rhône Alpes

Suite à l'approbation du premier Schéma Régional d'Organisation Sanitaire (SROS) en 1994 [25], les principaux établissements hospitaliers du secteur sanitaire n° 6 de la région Rhône-Alpes , le Centre Hospitalier Régional et Universitaire de Saint Etienne (CHRUSE) et l'Union Départementale de la Mutualité de la Loire (UDML) ont constitué un syndicat inter hospitalier nommé Institut de Cancérologie de la Loire (ICL) pour assurer la coordination des soins d'oncologie dans ce secteur sanitaire.

Par courrier en date du 06/11/97 [26] le directeur de l'ARH demandait à ces établissements d'« accompagner la mise en place de l'Institut de Cancérologie de la Loire » par la recherche de « l'élaboration de la file active de cancérologie » de chaque établissement et par l'étude de l'« inter-file active » entre ces établissements. Il était alors convenu de s'appuyer sur le PMSI, qui s'il ne constitue pas un dossier complet de cancérologie, permet de déterminer le nombre de personnes cancéreuses traitées par chaque établissement. Mais, en raison des contraintes d'anonymisation des données du PMSI imposées par la CNIL, se posait alors le problème de dénombrer les malades qui bénéficient d'une prise en charge partagée par ces établissements.

Le Service de Santé Publique et de l'Information Médicale (Pr RODRIGUES) du CHRUSE a appliqué le logiciel ANONYMAT aux noms, prénoms, dates de naissance de chacun des enregistrements des 3 bases de données issues des données

PMSI 1996 des 3 établissements concernés de façon à les rendre anonymes tout en pouvant les chaîner pour repérer les patients communs aux différents établissements [27]. De plus, la confrontation entre le nombre de numéros d'Anonymat obtenus dans chaque base et le nombre de patients calculé par l'administration a permis d'estimer le taux de doublons dans chacune des bases administratives.

5.2 Développement d'un recueil régional d'indicateurs en périnatalité en région Bourgogne

Un réseau périnatal s'est progressivement développé en Bourgogne depuis 1992 [28]. Ce réseau inclut tous les établissements prenant en charge les femmes enceintes et les nouveau-nés dans la région, soit à ce jour 20 établissements dont 4 privés. Les membres du réseau périnatal ont souhaité disposer d'une évaluation précise et continue de l'état de santé de la population concernée afin d'identifier les dysfonctionnements existants pour les corriger.

Un recueil régional continu d'indicateurs a été mis en place sur la base du volontariat pour toutes les naissances prises en charge dans les établissements de la région Bourgogne (environ 18 000 naissances annuelles). Un groupe de travail multidisciplinaire (pédiatres, obstétriciens, sage-femmes, médecins de santé publique, médecins de PMI, représentants des mutuelles, un directeur d'établissement) a retenu 42 indicateurs, 29 pour la mère et 13 pour l'enfant. Les informations sont extraites du PMSI, sous forme de Résumés d'Unité Médicale (RUM). Les indicateurs n'existant pas dans le PMSI (âge gestationnel, facteurs de risques psychosociaux) font l'objet d'un recueil supplémentaire sur une fiche adjointe au RUM, constituant un « RUM élargi ». Il a été nécessaire pour cela de modifier tous les logiciels de traitement des données PMSI des établissements participants. Sont collectés tous les « RUM élargis » correspondants aux séjours des mères (pour prise en charge de la grossesse, accouchement et suites de couches) ainsi qu'aux séjours des nouveau-nés en maternité et en services d'hospitalisation (pédiatrie, néonatalogie, réanimation néonatale). Ces RUM peuvent être recueillis dans plusieurs établissements différents pour une même mère ou un même nouveau-né (hospitalisations successives dans plusieurs établissements).

Pour le traitement des données médicales, le chaînage des « RUM élargis » est impératif et ceci à deux niveaux différents. D'une part, les « RUM élargis » d'une même personne, mère ou nouveau-né, doivent pouvoir être reliés lorsqu'il y a hospitalisations successives dans plusieurs unités, y compris lorsqu'il s'agit d'établissements différents. D'autre part, les « RUM élargis » de la mère doivent être reliés à ceux de ses enfants, même si ceux – ci sont hospitalisés dans un établissement différent, afin d'évaluer l'impact postnatal des facteurs de risques et des pathologies gravidiques. Toutefois, conformément à la législation, les fichiers ne sont transmis au DIM du CHU de Dijon pour exploitation qu'après avoir été rendus anonymes. Le chaînage de données anonymes a alors été rendu possible par

l'utilisation du logiciel ANONYMAT, à partir de six variables : le nom de jeune fille de la mère, son prénom et sa date de naissance, le prénom de l'enfant et sa date de naissance, le code postal de résidence de la mère. Ces six informations sont saisies de manière identique dans les « RUM élargis » de la mère et de son bébé. Dans le cas de grossesse multiple, tous les prénoms des nouveau-nés doivent être saisis dans le « RUM élargi » de la maman. Ces six variables nominatives sont utilisées par le programme de chaînage, après avoir été rendues anonymes.

A ce jour, 18 établissements sur 20 effectuent le recueil d'indicateurs en routine, représentant 95 % des naissances de Bourgogne. Les 2 établissements restants se sont engagés à débuter le recueil en septembre 2000. Avant transmission, les fichiers sont validés au sein de chaque établissement par comparaison avec les cahiers de services (maternités et services de pédiatrie). De plus, l'exhaustivité et la qualité du recueil des données de chaînage sont systématiquement contrôlées dans chaque établissement et de façon centralisée par l'équipe coordinatrice (Dr CORNET) qui traite les données au DIM du CHU (tests de chaînage mère-enfant, identification des fiches non chaînées, correction des erreurs). Un chaînage mère – enfant est obtenu pour 72% des nouveau-nés, avant validation, et pour 100% des nouveau-nés après l'ensemble des procédures informatisées et manuelles de correction des erreurs.

6. Conclusion

L'intérêt de la solution proposée dans cet article est d'assurer un anonymat irréversible de l'identité, tout en permettant le chaînage des informations, sans être un obstacle à la validation de ces données. Chaque source de l'information (par exemple un établissement) est en droit de conserver la correspondance entre le numéro d'anonymat et l'identité du patient, et peut donc procéder aux vérifications et corrections demandées par le coordinateur du réseau.

REFERENCES

1- Quantin C, Allaert FA, d'Athis P, Dusserre L. Can a database be anonymous? MIE 99, Slovenia, 22-26 août 1999:297-301.

2- Loi 98-1266 du 30 décembre 1998 (article 107). Loi de finances pour l'année 1999.

3- Décret définissant les conditions dans lesquelles sont souscrites les déclarations et accordées les autorisations concernant les moyens et prestations de cryptologie, n° 98-101 du 24 février 1998.

4- Décret fixant la liste des moyens et des prestations de cryptologie dispensées de toute formalité préalable, n° 98-206 du 23 mars 1998.

5- Décret fixant la liste des moyens et des prestations de cryptologie pour lesquels la déclaration se substitue à l'autorisation, n° 98-207 du 23 mars 1998.

6- Décret n°99-199 du 17 mars 1999 définissant les catégories de moyens et de prestations de cryptologie pour lesquelles la procédure de déclaration préalable est substituée à celle d'autorisation.

7- Fisher F, Madge B. Data security and patient confidentiality : the manager's role. *Int J Biomed Comput* 1996;43:115-119.

8- Douglas S. Cryptologie, théorie et pratique,1996, International Thomson Publishing.

9- Beckett B. Introduction aux méthodes de cryptologie, 1990, Masson.

10- Brassard G. Cryptologie contemporaine, 1993, Masson.

11- Dusserre L., Quantin C., Bouzelat H.A one way public key cryptosystem for the linkage of nominal files in epidemiological studies. MEDINFO 95, R.A. GREENES, H.E. PETERSON, D.J. PROTTI (editors), *Elsevier Science Publishers* (North-Holland):644-647.

12- Quantin C, Bouzelat H, Allaërt FA et al. How to ensure data security of an epidemiological follow-up : quality assessment of an anonymous record linkage procedure. *Int J Med Inf* 1998;49:117-22.

13- Michaelis J, Miller M, Pommerening K et al. A new concept to ensure data privacy and data security in cancer registries. *Medinfo* 1995;8:661-5.

14- Loi 90-1170 du 29 décembre 1990 sur la réglementation des télécommunications.

15- Loi 96-659 du 26 juillet 1996 sur la réglementation des télécommunications (modifiant la loi 90-1170 du 29 décembre 1990.

16- Trouessin G, Allaërt FA. FOIN : a nominative information occultation function. MIE 97;43:196-200.

17- Trouessin G. Rapport " qualité diagnostique et thérapeutique en cancérologie : communication d'informations multimédia dans un réseau sécurisé multidisciplinaire. Sécurité de l'information médicale en télémédecine ", étude du ministère de la recherche.

18- Quantin C., Bouzelat H., Allaërt FA et al. Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Meth Inform Med* 1998;37:271-7.

19- Brassard G. Modern Cryptology, 1993, Lecture Notes in Computer Science, 1993.

20- Schneier B. Applied Cryptography, Protocols, Algorithms, and Source Code in C. John Wiley & Sons Inc, 1994.

21- Brenner H. Schmidtmann I, Stegmaier C. Effects of record linkage errors on registry-based follow-up studies. *Stat Med* 1997;16(23):2633-43.

22- Bouzelat H. Anonymat et chaînage de fichiers médicaux en vue d'études épidémiologiques. Thèse de Docteur d'Université spécialiste en Informatique Médicale. Université de Bourgogne. 1998:p. 97.

23- Jaro M.A. Probabilistic-linkage of large public health data files. *Statistics in Medicine* 1995;14:491-8.

24- Sugarman JR, Holliday M, Ross A et al. Improving American Indian cancer data in the Washington state cancer registry using linkages with the Indian health service and tribal records. *American Cancer Society* 1996;78(7suppl):1564-8.

25- Abrial V. Les contrats d'objectifs entre les établissements publics de santé et l'agence régionale de l'hospitalisation : analyse d'environnement du CHU de St Etienne. *Thèse de Docteur en Médecine.* Université de Franche-Comté. 1998.

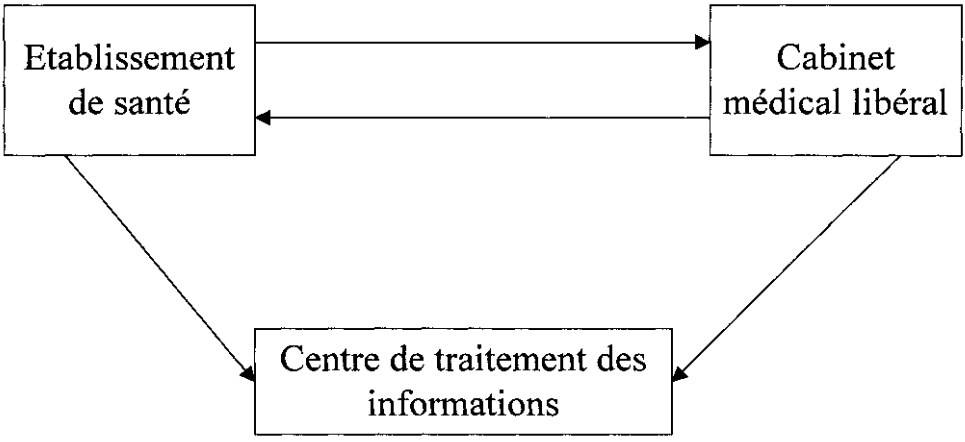
26- Agence Régionale de l'Hospitalisation de Rhône-Alpes. Mission d'enquête sur les dépenses médicales et pharmaceutiques : Lyon 6 novembre 1997.

27- Quantin C, Allaert FA, Bouzelat H, Rodrigues JM, Trombert-Paviot B, Brunet-Lecomte P, Gremy F, Dusserre L. La sécurité des réseaux d'informations médicales : application aux études épidémiologiques. *Revue d'Epidémiologie et de Santé Publique*. 2000;48:89-99.

28- Cornet B, Métral P, Fromaget J, Sagot P, Gouyon JB. Réseau périnatal de Bourgogne. Technologie et Santé, 1999;37:51-56.

Figure 1 : **Intérêt des méthodes de cryptage pour assurer la sécurité des informations**

Confidentialité = Cryptage des flux



Anonymat ⇨ transformation irréversible de l'identité (hachage)

Figure 2 : Utilisation du logiciel Anonymat et des clés dans le cadre d'une étude épidémiologique multicentrique