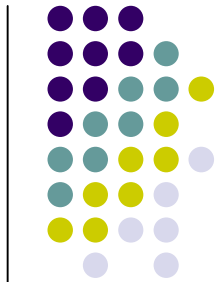
Le chaînage de bases de données anonymisées pour les études épidémiologiques multicentriques nationales et internationales : proposition d'un algorithme cryptographique

Catherine Quantin[1], Maniane Fassa[1], Gilles Trouessin[2]

1- Inserm, U866, Dijon, F-21000, France ; Univ de Bourgogne, Dijon, F-21000, France ; CHRU Dijon, Service de Biostatistique et d'Informatique Médicale, Dijon, F-21000, France(CHU Dijon)

2- . OPPIDA Sud, Toulouse



Rappel sur l’Anonymisation et le Chaînage des Données

Procédure d'Anonymat

HACHAGE A SENS UNIQUE

➔ **Transformation irréversible de l'identité (algorithme Standard Hash Algorithm – SHA)**

➔ **Code identique et permanent, pour une identité donnée, pour permettre le regroupement des informations d'un même individu.**

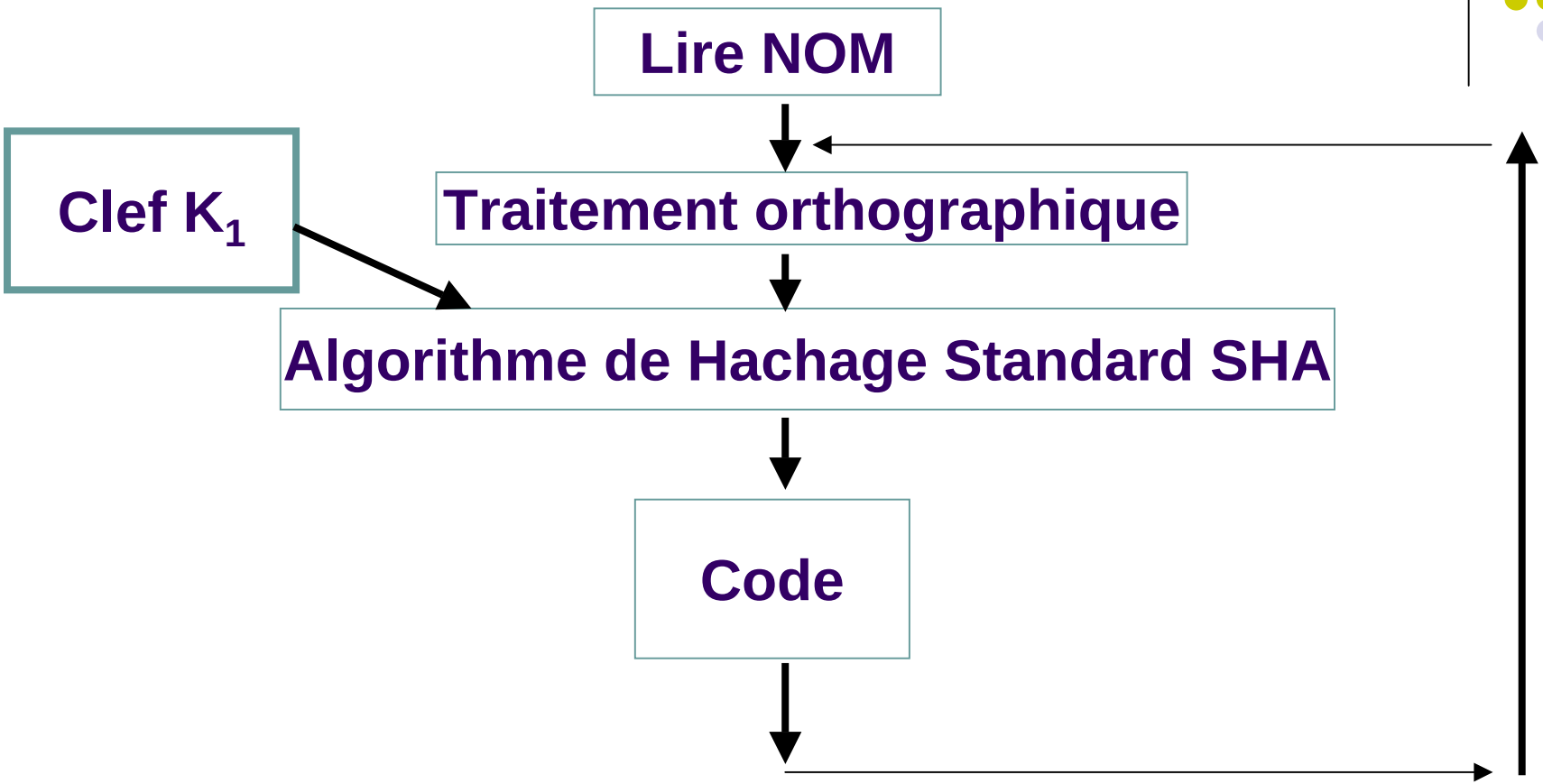
Méthode d'Anonymat

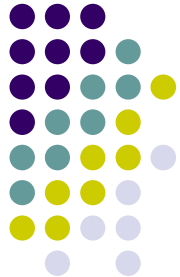
DECLAREE AUPRES

. du Service Central de la Sécurité
des Systèmes d'Information (SCSSI) : mars 1996

. de la Commission Nationale de l'Informatique et des Libertés (CNIL) : décembre 1996

Algorithme d'Anonymat





Chaînage des données

Après hachage des données d'identification plusieurs méthodes de chaînage sont possibles :

1 - Méthodes déterministes, dont le chaînage direct après hachage (méthode FOIN)

2 - Méthodes probabilistes

Applications

Pour la région Bourgogne : réseau périnatal, réseau hépatite C

Pour la région Franche-Comté : réseau hépatite C, tentative de suicide

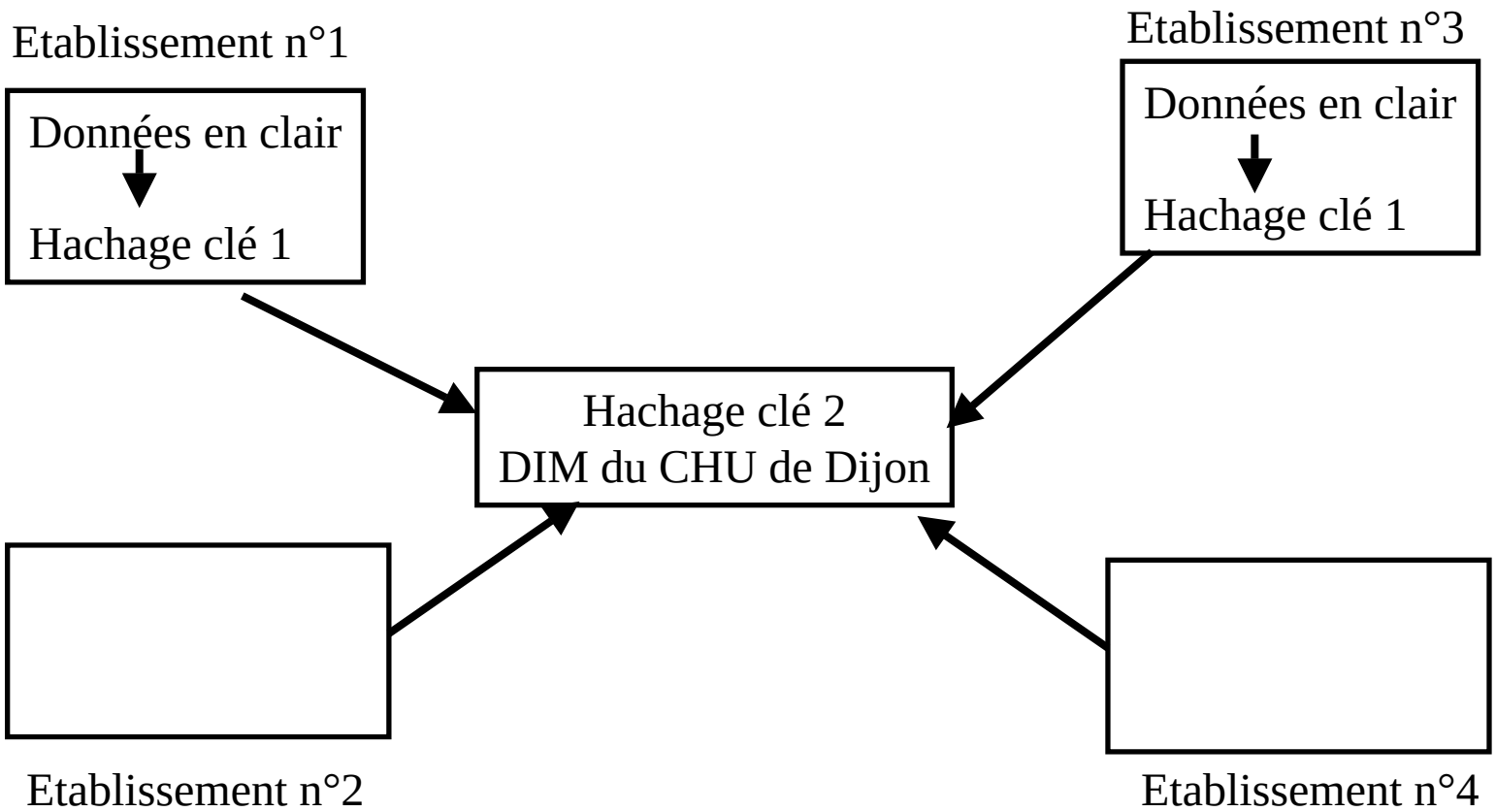
Pour la région Rhônes-Alpes : suivi de la file active de cancérologie dans le département de la Loire

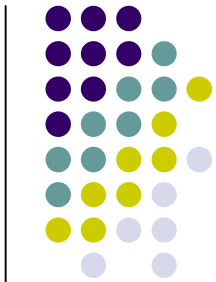
Au niveau national : données du PMSI (logiciel FOIN-CESSI/CNAMTS), suivi des notifications d'infection par le virus du SIDA (Institut de Veille Sanitaire), suivi des Rmistes, suivi des étudiants et des élèves (MENRT)

Au niveau international : suivi des patients de Suisse et pour les registres du Luxembourg

Un exemple d'application

- Réseau Périnatal de Bourgogne

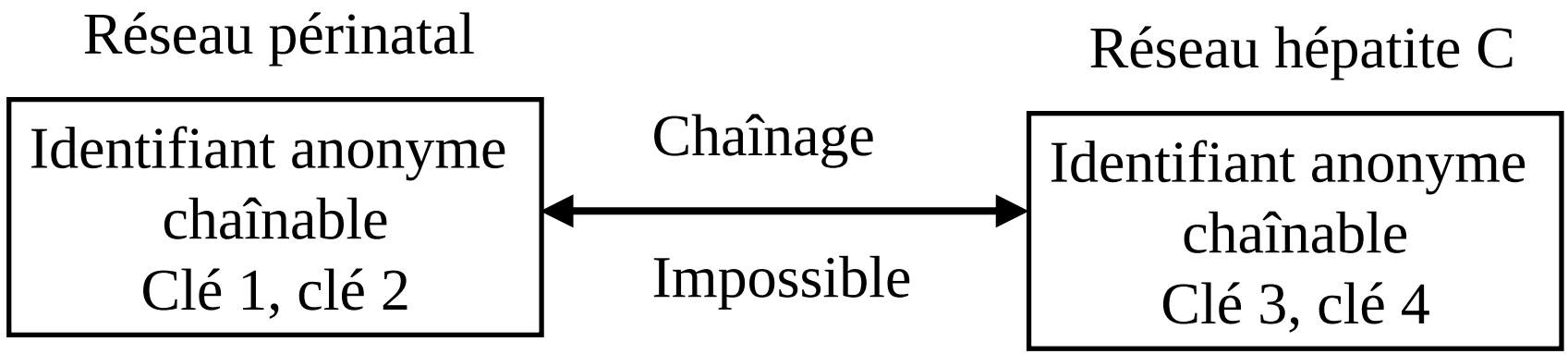




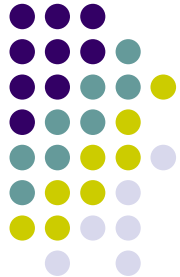
- Il est recommandé d’utiliser des clés différentes pour chaque étude

- Cette recommandation est en contradiction avec les besoins ultérieurs de chaînage

Exemple de besoin de chaînage



Nécessité d'une nouvelle méthode



- rendre techniquement possible des appariements sécurisés, alors que ceux-ci n'avaient initialement pas été prévus,

  ➔ permettre le chaînage des données entre bases de données, (études multi-centriques), dans les conditions de sécurité requises par la loi et par la CNIL

Méthode

- Le double hachage est réalisé par des clés communes à toutes les études
- La sécurisation de l'identifiant doublement haché

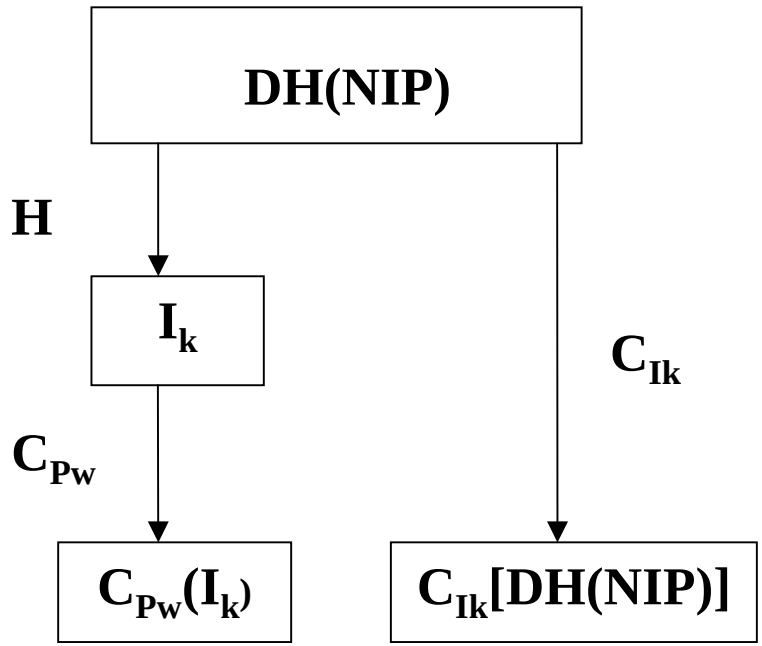
DH (NIP) est réalisé par une fonction de chiffrement

- La clé utilisée pour le chiffrement va dépendre de la valeur de l'identifiant doublement haché DH (NIP)

  ➔ pour chaque identifiant, une clé différente Ik sera utilisée
- Cette clé Ik est sécurisée par chiffrement avec le mot de passe de l'étude (Pw)

NIP : Numéro d'Identifiant du Patient

Sécurisation par chiffrement de l'identifiant haché



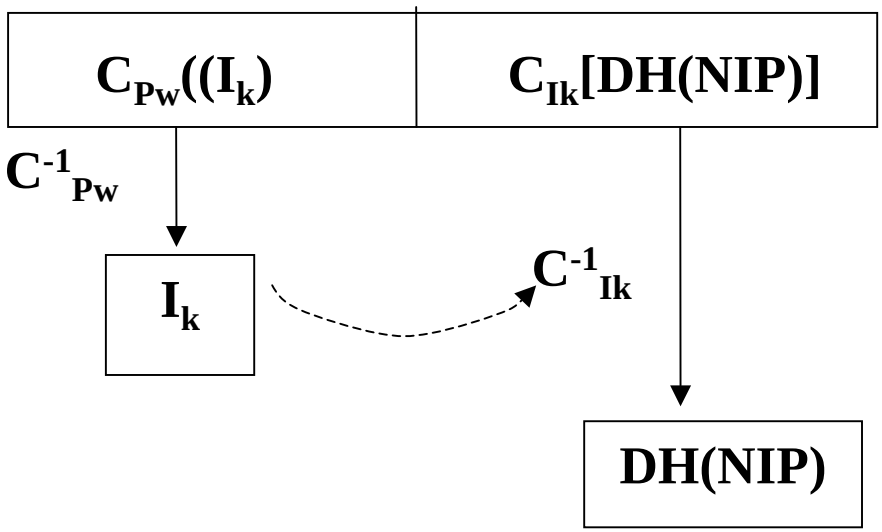
H = fonction de Hachage ; C = fonction de Chiffrement ; DH : Double fonction de Hachage

Ik : clé variable dépendant de l'identité du patient ; PW : clé unique définie pour l'étude

Pour déchiffrer l’identifiant DH (NIP) il faut :

- Retrouver la clé Ik, ce qui nécessite d’avoir le mot de passe de l’étude
- Déchiffrer l’identifiant haché grâce à cette clé Ik

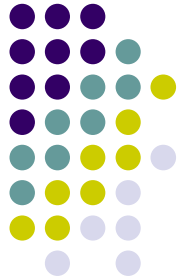
Déchiffrement de l'identifiant haché

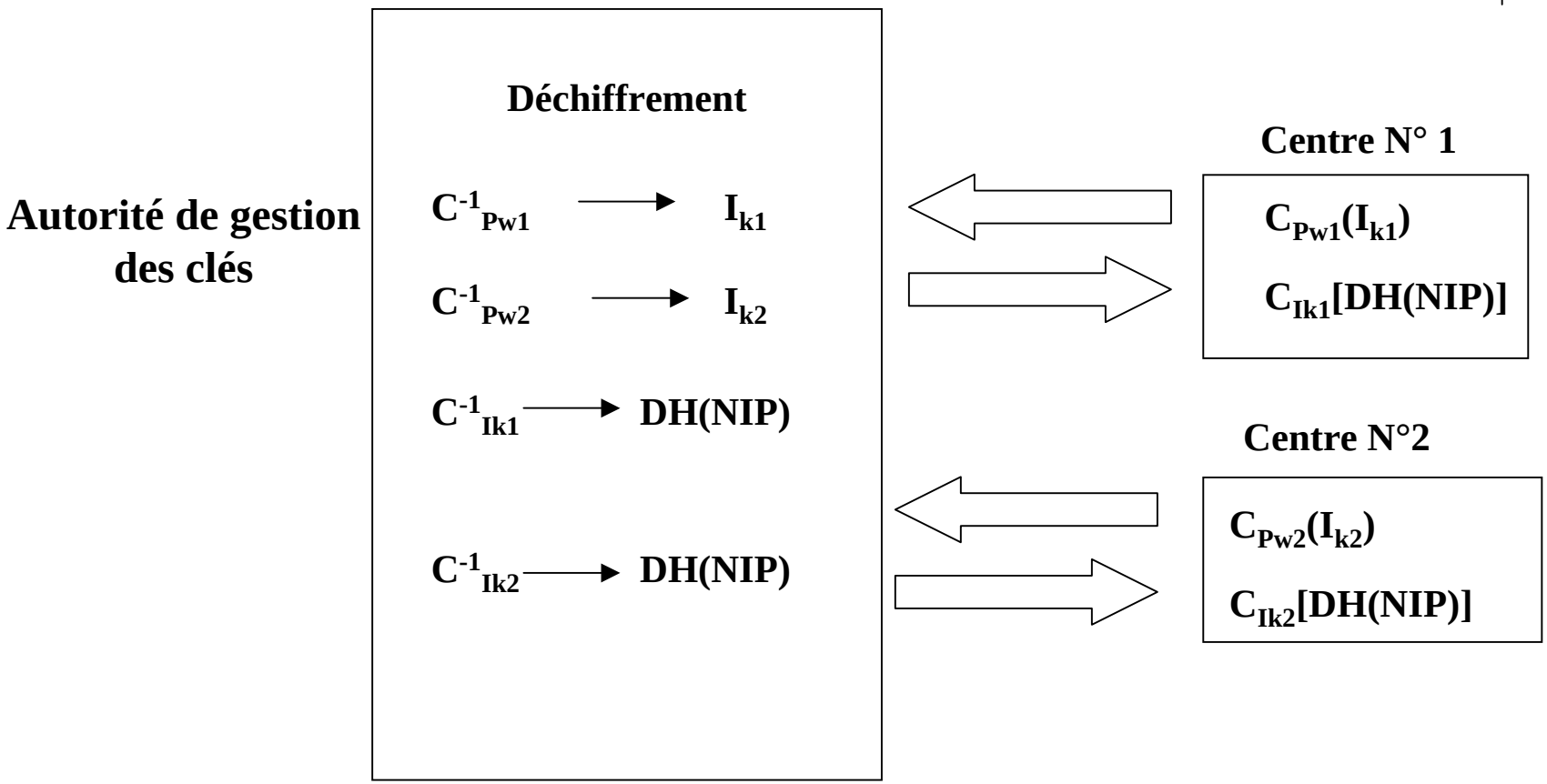


- C^{-1} = fonction de Déchiffrement (associée à C) ;
  : la clé Ik est utilisée pour la fonction C^{-1}

- Il est donc possible de réaliser un chaînage inter-centres, si une autorité de gestion des clés :
  - reçoit les données de chaque centre,
  - déchiffre les clés Ik_1 et Ik_2, grâce aux mots de passe Pw_1 et Pw_2 de chaque étude,
  - déchiffre les identifiants hachés, grâce aux clés Ik_1 et Ik_2,
  - relie les observations se rapportant au même identifiant haché,
  - retourne les fichiers appariés aux centres.

Chaînage inter-centres à partir de l'identifiant haché, par l'Autorité de Gestion des clés





Conclusion

- Il est donc possible de relier, en toute sécurité, des données provenant d'études différentes, à condition que :

  - Le même algorithme de hachage, soit utilisé, avec les mêmes clés, pour chaque étude
  - Une autorité de gestion des clés assure le rapprochement
    - ➔ Qui pourrait assurer ce rôle ?
    - ➔ Faut il créer un CENTRE D'APPARIEMENTS SECURISES ?

Références

- QUANTIN C., BOUZELAT H., ALLAERT F.A., BENHAMICHE A.M., FAIVRE J., DUSSERRE L. How to ensure data security of an epidemiological follow-up : quality assessment of an anonymous record linkage procedure, International Journal of Medical Informatics. 1998; 49:117-122.
- QUANTIN C., ALLAERT F.A., DUSSERRE L. Anonymous statistical methods versus cryptographic methods in epidemiology. International Journal of Medical Informatics 2000;60:177-83.
- QUANTIN C, BINQUET C, ALLAERT FA, CORNET B, PATTISINA R, LE TEUFF G, FERDYNUS C, GOUYON JB. Decision analysis for the assessment of a record linkage procedure : application to a perinatal network. Methods of Information in Medicine, 2005;44 (1) :72-9.
- QUANTIN C, ALLAERT FA, GOUYON B, COHEN O. Proposal for the creation of a European healthcare identifier. Stud Health Technol Inform, 2005;116:949-54.
- QUANTIN C, COHEN O, RIANDEY B, ALLAERT FA. Unique patient concept: a key choice for European epidemiology. International Journal of Medical Informatics, 2007;76:419-426.
- FOURNEL I, SCHWARZINGER M, BENZENINE E, BINQUET C, HILL C, QUANTIN C. Détermination du statut vital par chaînage entre des données hospitalières et les données de mortalité nationales anonymisées. Revue d'Epidémiologie et de Santé Publique 2007;55:365-373.
- QUANTIN C, ALLAERT FA, FASSA M, AVILLACH P, COHEN O. Making Health Identification Interoperable in Europe. Hospital IT Europe special issue, 2007:47-48.

Références

- QUANTIN C, ALLAERT FA, FASSA M, RIANDEY B, AVILLACH P., COHEN O. How to manage a secure direct access of European patients to their computerised medical record and personal medical record. Stud health Technol Inform 2007;127:246-55.
- QUANTIN C., ALLAERT F., FASSA M., AVILLACH P., FIESCHI M., COHEN O. Interoperability issues regarding patient identification in Europe. Conf Proc IEEE Eng Med Biol Soc, 2007; 2007 : 6161.
- QUANTIN C., ALLAERT F., AVILLACH P., RIANDEY B., FIESCHI M., FASSA M, COHEN O. Proposal of a french health identification number interoperable at the European level. Medinfo, 2007, 12 : 503-7.
- QUANTIN C, ALLAERT FA, AVILLACH P, FASSA M, RIANDEY B, TROUESSIN G, COHEN O. Building application-related patient identifiers: what solution for a European country? International Journal of Telemedicine and Application 2008, Article ID 678302, 1-5.
- QUANTIN C, FASSA M, COATRIEUX G, RIANDEY B, TROUESSIN G, ALLAERT FA. Chaînage de bases de données anonymisées pour les études épidémiologiques multicentriques nationales et internationales : proposition d'un algorithme cryptographique. Accepté dans la Revue d'Epidémiologie et de Santé Publique.
- QUANTIN C, FASSA M, COATRIEUX G, TROUESSIN G, ALLAERT FA. Combining hashing and enciphering algorithms for epidemiological analysis of gathered data. Methods of Information in Medicine 2008;5:454-458.

Contacts

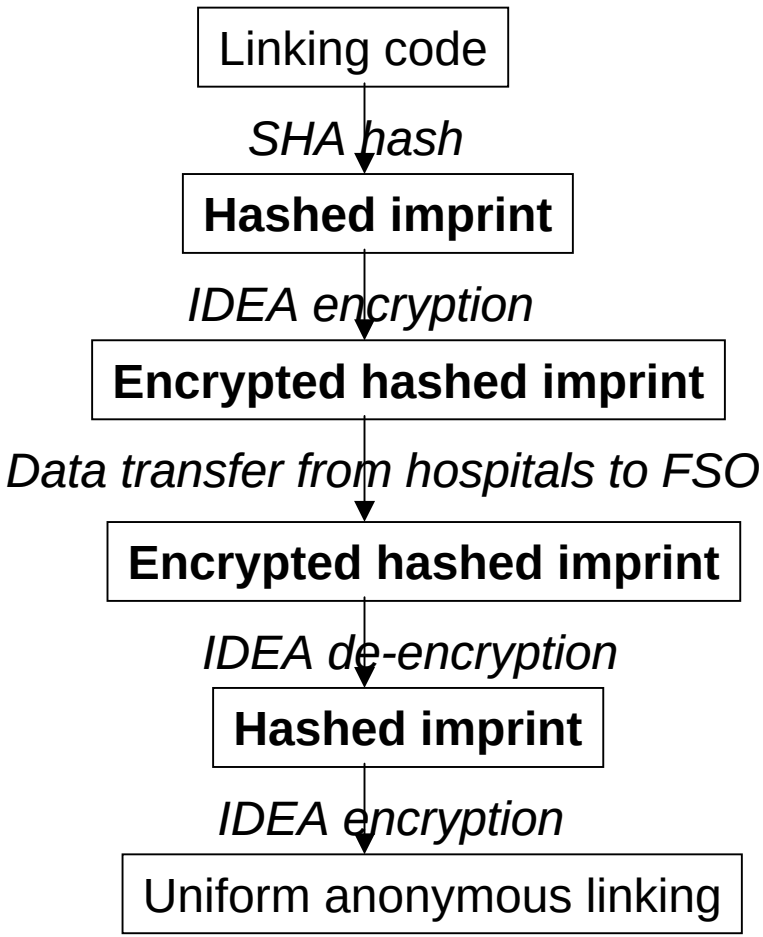
- HC Forum : Olivier COHEN – olivier.cohen@hcforum.fr
- OPPIDA : Gilles TROUESSIN – gilles.trouessin@oppida.fr
- CHU de Dijon : Catherine QUANTIN – catherine.quantin@chu-dijon.fr

Annexe

- La différence avec l'algorithme mis en place en Suisse[1] vient du choix de la clé de chiffrement : ici la clé Ik est variable en fonction
  - De l'étude
  - De l'identité du patient
- Ceci permet de faire varier la clé en fonction de l'étude tout en permettant le rapprochement des données d'études différentes

1- BORST F., ALLAERT F-A., QUANTIN C. The swiss solution for anonymously chaining patient files. Proc. MEDINFO 2001 ; IMIA 2001 : 1239-41.

Figure 1 – Organizational procedure, article BORST, ALLAERT, QUANTIN



IDEA : International Data Encryption Algorithm
FSO : Federal Statistical Office

Validation des données pour le rapprochement

Après hachage :

- On peut demander la correction des données correspondant à un numéro d'anonymat.

- Les informations corrigées par la source sont retransmises, après nouveau hachage.