Square des Utilisateurs

Sécurité(s) et Intimité…

…des données à caractère personnel

L'article paru dans la lettre d'ADELI n°42 de janvier 2001, avait initialisé une réflexion sur la sécurité en en proposant de lever les ambiguïtés dues, entre autres, à des synonymies entre Sécurité(s) & Sûreté(s) au travers d'une présentation générale de la Sécurité, de la Sûreté et de la Sûreté de Fonctionnement des systèmes d'information. Cet article évoquait l'intimité-"privacy" qui protège la vie privée en termes assez novateurs de confidentialité-séclusion[1] par opposition à la confidentialité-discrétion[2] habituelle qui préserve le secret en des termes plus classiques et usuels. Manipuler des données à caractère personnel, pose la problématique du respect de la vie privée des individus et citoyens. Quelles spécificités peut-on, ou doit-on, ajouter à celles habituellement exprimées en sécurité des systèmes d'information ? En particulier qu'entend-t-on par :

- *…respect de la vie privée au sens habituel du terme ?* *S'agit-il d'intimité ?*
- *…par protection des données personnelles électroniques ?* *S'agit-il de 'e-privacy' ?*
- *…par garantie de l'anonymat le plus rigoureux possible ?* *S'agit-t-il de 'séclusion' ?*

Cet article présente quelques notions sémantiques et terminologiques relatives à la sécurité, mais aussi et surtout, à l'intimité des données à caractère personnel. Telles sont, à titre illustratif mais très significatif, les données personnelles de santé/soins/médicales/hospitalières. Cet article apporte des éléments qui montrent que la prise en compte de l'intimité-'privacy' crée un métier, sinon nouveau, du moins émergent parmi ceux plus classiques et traditionnels de la sécurité informatique.

Nos exemples sont issus de la sécurité des systèmes d'information de santé, pratiquée plus de huit ans dans l'Assurance Maladie. Ce sont des besoins et de solutions qui pourront être développés et adaptés dans des secteurs connexes : médical, social, socio-sanitaire, vigilance et veille sanitaires. Ces exemples pourront être étendus à toutes les démarches que les citoyens ou communautés d'intérêts pourront mener via les nouvelles technologies de l'information et de la communication : consultations, sondages et votes électroniques, forums de discussions sur internet/web, …

Nous ne traiterons pas les aspects purement juridiques de la sécurité et intimité des données à caractère personnel, sous-jacents à cette problématique. Il s'agit d'un métier à part entière, complexe, très précis et très rigoureux. Nous ne traiterons pas, non plus, les détails des services, techniques et outils destinés à faire respecter l'intimité-'privacy' des données personnelles électroniques.

Introduction

Voici quelques mots-clés et expressions propres au sujet abordé que nous vous proposons d'examiner au cours de cette brève présentation.

Un dilemme, en particulier pour les systèmes d'information de santé et de soins, provient du fait qu'il faut souvent allier les deux sécurités : sécurité-'*security*' et sécurité-'*safety*'. Faut-il privilégier l'une au détriment de l'autre ? Ou, au contraire, faut-il rechercher l'équilibre ?

---

[1] *'séclusion' : mot d'origine latine (**seclusum**, **secludere** : séparer de). Adaptation physiologique par laquelle un animal ou une plante s'isole du milieu, empêchant passivement les actions défavorables de s'exercer sur lui. [Le Robert]*

[2] *'discrétion' : du bas latin (**discretio**, **-onis** : discernement). Attitude de quelqu'un qui sait garder un secret. A la discrétion de quelqu'un : à sa disposition, à sa merci, en son pouvoir. [Le Robert]*

Un élément de complication de système manipulant des données à caractère personnel provient aussi de la cohabitation entre qualité et intégrité : *qualité* des processus versus *intégrité* des données. Faut-il investir dans la qualité, voire la certification, des processus, au risque de négliger les données ?

Le manque frappant de distinction entre les formes de confidentialité à respecter pour préserver l'intimité électronique pose un problème : doit-on se contenter de respecter la confidentialité par la *discrétion* (tant que celle-ci ne sera pas levée, légitimement ou non) ou peut-on obliger de la respecter par la *'séclusion'* (c'est-à-dire dès que celle-ci est instaurée, irrémédiablement) ?

Parmi les outils au service de la confidentialité, il faut rappeler la cryptographie, ou art des écritures secrètes, dont des applications récentes par rapport au classique chiffrement/déchiffrement sont : l'*occultation* des noms, l'*anonymisation* des données, la *pseudonymisation* des informations identifiantes. Pourquoi, quand, comment occulter plutôt qu'anonymiser ou pseudonymiser ?

En effet, une démarche d'occultation, d'anonymisation, de pseudonymisation peut vite devenir complexe et tout aussi sûrement inefficace si ne sont pas posées à temps les bonnes questions d'expressions de *besoins*, de réalisation d'*objectifs* et de formalisation d'*exigences*. Occultation, anonymisation et pseudonymisation pourront être *réversibles* (abus de langage), *irréversibles* (d'une manière générale), voire *inversibles* (dans certaines situations précises). Comment décider ?

On pourra parfois répondre à des exigences de *chaînages* (entre données anonymisées) mais aussi à des exigences de *robustesse* (des identifiants anonymisés ou pseudonymisés). Que faut-il choisir ?

Exigences de chaînage : si les techniques de *fragmentation* des fichiers et de données répondent utilement à des soucis de type *confidentialité-discrétion*, la technique de *segmentation* des liens de chaînage entre données anonymes répond parfaitement à un souci de *confidentialité-séclusion*, tout en respectant, précisément, les exigences de chaînage acceptables et légitimes.

Exigences de robustesse : ces exigences doivent faire face aux risques de *désanonymisation directe* (cryptographique), *indirecte* (procédurale) ou par *inférence* (en logique) d'informations nominatives à partir de données non nominatives, en cas de trop grande corrélation d'informations.

Sécurité(s) des données à caractère personnel : développements

Rappel du trio sécurité : aspects juridiques-procéduraux-techniques

Dans la lettre n°44 d'ADELI de janvier dernier, nous avons introduit le trio de la sécurité :

- *responsabilité* des acteurs ou composante *juridique* ;
  (qui est responsable de, et pourquoi, mettre en oeuvre la sécurité ?) ;
- *superstructure* organisationnelle ou composante *organique* ;
  (où, quand, faut-il, peut-on, doit-on sécuriser le système ?) ;
- *infrastructure* opérationnelle ou composante *technique* ;
  (comment faire et faire en sorte de mettre en œuvre la sécurité ?).

Que ce soit en sécurité-*'security'* ou en sécurité-*'safety'* :

- des aspects juridiques orientent tout d'abord les décisions et les actions, telles que, l'obligation de confidentialité et d'intégrité des données à caractère personnel, faite par la loi Informatique et Libertés ou la directive européenne sur la protection des données ;
- puis des aspects organisationnels mettent ensuite ces décisions en applications ; par exemple, la séparation des rôles et des pouvoirs et habilitations interdisant la rencontre, dans des systèmes médico-financiers ou médico-administratifs, entre d'un côté des données purement médicales, de type diagnostics ou pathologies, et d'un autre côté l'identifiant national des personnes servant de numéro dit, abusivement, « de sécu. » ;
- et enfin des techniques, telles que le contrôle d'accès aux informations ou la cryptographie dans les échanges électroniques, les réalisent concrètement.

Le dilemme Sécurité-Sûreté en santé

Deux formes de sécurité sont souvent amenées à cohabiter, en particulier dans des systèmes ou applicatifs critiques manipulant des données de santé, donc à caractère personnel :

- *Sécurité-"security" :* dans la mesure où les données personnelles manipulées doivent être rendues suffisamment disponibles, intègres et confidentielles pour être considérée comme 'de confiance', ou authentiquement crédibles, mais aussi être parfaitement auditables ;
- *Sécurité-"safety" :* dans le sens où certaines données, parfois anodines telles que celles relevées et rendues accessibles par une infirmière ou une aide-soignante, sont cruciales, en termes de pertinence et de fiabilité pour la santé, voire l'innocuité, du patient ou du malade.

La '*security*' répond aux exigences fonctionnelles et apporte les solutions *ad hoc* comme les politiques d'autorisation ou interdiction, à accéder, modifier, lire des informations. La '*safety*' impose des contraintes particulières et procédures spécifiques, comme le partage d'information, pour donner pleine confiance dans la réactivité et la sûreté du système de soins et de ces différents processus.

Ainsi, on doit trouver un juste milieu entre trop de confidentialité au prétexte du respect de la vie privée et trop peu de confidentialité pour des soucis de partage efficient d'information. Ici se place utilement le débat entre confidentialité-*discrétion* et confidentialité-*séclusion*, et aussi entre '*security*' et '*safety*' d'une manière générique.

Sécurité(s) par la disponibilité

À quel niveau doit-on placer l'exigence de disponibilité de l'information de santé ou de soins, dans le contexte des établissements/structures délivrant des soins ou devant prendre des décisions vitales ?

Trop d'exigences de disponibilité coûte très cher et ampute d'autant les efforts consentis sur les autres aspects de la sécurité ou sur d'autres composantes du système d'information. Mais, c'est parfois la donnée la plus anodine qui peut permettre de construire un diagnostic critique ou sérieux. On souhaiterait donc disposer, toujours et partout, de l'information utile et pertinente.

À l'inverse, trop peu d'exigences de disponibilité, paralyse un système insuffisamment opérant, voire totalement inopérant. Le système d'information se réduit alors à un simple système informatique, incapable de fournir en temps utile les éléments d'information décisifs et/ou les processus de décision, eux-mêmes. Dès lors qu'il s'agit de système vital, ce n'est pas acceptable.

Un juste équilibre de la politique de disponibilité et une bonne opérationnalité de celle-ci contribuent à la fois à la sécurité-'*security*' et à la sécurité-'*safety*'.

Sécurité(s) par l'intégrité

À quel niveau placer l'exigence d'intégrité des données de santé ou de soins, dans le contexte des établissements et structures des soins. Les systèmes d'aide aux décisions autour des politiques socio-sanitaires, dans les caisses maladie, vieillesse, famille (nationales, régionales, voire locales) ont aussi besoin d'information suffisamment intègres, implicitement précises et correctes, pour travailler en toute pertinence, mais au détriment parfois du respect de la vie privée des assurés sociaux et des malades.

En effet, trop d'exigences d'intégrité ont pour incidence, par exemple, de gérer une masse faramineuse de détails informationnels, sous prétexte de pouvoir en avoir besoin le moment venu. Rappelons que c'est parfois la donnée la plus anodine qui peut permettre de construire une démarche sanitaire ou sociale pertinente car équitable et juste à temps. On entre dans des débats du style : précision *versus* pertinence, correction *versus* fraîcheur, qualité *versus* fiabilité des données.

A l'inverse, un système conçu avec trop peu d'exigences d'intégrité serait de mauvaise qualité et peu efficace, ce qui est inacceptable lorsque sont manipulées des données médicales vitales ou critiques pour la santé des patients, ou critiques ou sensibles pour l'estimation de l'état sanitaire d'une catégorie de population, ou pour un espace-temps ou un périmètre géographique et/ou thématique donné.

Un juste équilibre de la politique d'intégrité et une bonne efficacité de celle-ci contribuent à la fois à la sécurité-'*security*' et à la sécurité-'*safety*'.

Sécurité(s) par la confidentialité

Pourquoi et comment gérer la confidentialité dès que l'on touche à des données à caractère personnel, notamment des données de santé ? L'Ordre National des Médecins rappelle que toute donnée de santé, dès lors qu'elle est connue d'un médecin, devient secret médical. Le secret médical est un secret professionnel « d'ordre public », ce qui veut dire que même le malade ou le patient ne peut délier son propre médecin du secret médical le concernant.

Il faut être très prudent et professionnel pour conseiller, décider et mettre en oeuvre des solutions de sécurité par la confidentialité. Une subtilité s'impose désormais pour distinguer deux notions appelées respectivement confidentialité-*discrétion* et confidentialité-*séclusion* :

- *discrétion*[3] *:* ce terme désigne la notion de ce qui est confidentiel, bien évidemment, et secret. Le secret en question, et par suite la discrétion y afférant, pourrait être levé dans des conditions qui doivent être précisées en vertu du trio sécurité déjà évoqué : conditions juridiques (légales, réglementaires), organisationnelles (par qui, sur ordre de qui ?), techniques (selon quelle méthode, protocole, technique, outils, algorithme ?) ;
- *séclusion*[4] *:* ce terme désigne une notion qui traduit la volonté de rester reclus dans son intimité (électronique, en l'occurrence). Forme de réclusion active, par choix délibéré et non pas par force (comme le laisserait entendre le terme *'réclusion'*) la *'séclusion'* est un droit de chacun en vertu, là encore, de conditions issues du trio sécurité évoqué *supra* : conditions juridiques (cf. loi Informatique et Libertés de 1978, directive européenne sur la protection des données personnelles de 1995, …), organisationnelles (quand et où garantir la vie privée des individus, en version 'données personnelles électroniques'), techniques (comment et avec quel degré de confiance et de robustesse garantir un anonymat, par exemple).

Par analogie avec les arguments formulés dans les deux paragraphes précédents, *'sécurité par la disponibilité'* et *'sécurité par l'intégrité'*, trop ou pas assez de confidentialité serait nuisible au système ; trop ou trop peu de confidentialité conduit à paralyser l'utilité du système ou de l'applicatif :

- par exemple, interdire tout accès aux données nominatives à des utilisateurs dont le métier est de manipuler de telles données pour rendre un service à l'usager directement, serait inutile et inapproprié ; de même, utiliser la cryptographie coûteuse en temps de calcul (voir ci-dessous chiffrement asymétrique) pour échanger de gros volumes de données, serait paralysant et tout aussi inapproprié ;
- à l'inverse, accorder un accès universel à l'information nominative par manque de discernement aboutirait à une perte de confiance totale dans le système, à son rejet et à sa non-utilisation. Refuser tout recours à de la cryptographie, dont le chiffrement symétrique, au seul motif que c'est une technologie intrinsèquement complexe ou *a priori* coûteuse, sans plus analyser les besoins, serait un manque de sérieux manifeste et conduirait à une méfiance compréhensible.

On découvre, ici aussi, qu'une mauvaise option de confidentialité serait nuisible au système et à la confiance qu'il inspire. Faire appel à la confidentialité-discrétion en lieu et place de la confidentialité-séclusion ne permettrait pas de répondre à une attente de respect de la vie privée ou, inversement, mettre en œuvre la confidentialité-séclusion en lieu et place d'une confidentialité-discrétion ne permettrait pas au système répondre aux fonctionnalités attendues :

- par exemple, chiffrer les noms des patients, dans une application d'accès à des bases de données statistiques sur les consommations de soins permettrait de masquer de façon seulement transitoire, c'est-à-dire 'occulter', les noms des patients. Un utilisateur de cet applicatif, ayant accès à la fonctionnalité de déchiffrement, pourrait déchiffrer les noms des patients, alors que son métier de statisticien ne le légitime pas pour cela. Il s'agit avant tout de garantir la

---

[3] La discrétion, dont une traduction pertinente en anglais est 'secrecy' (ce qui est secret), correspond à la cryptographie dite 'à clé secrète', car cette clé secrete de chiffrement est partagée entre l'émetteur et le récepteur d'un message (en général).

[4] La séclusion, dont la traduction immédiate en anglais est 'to seclude' et une traduction plus indirecte, mais en cohérence avec la discrétion-'secrecy', sera plutôt 'privacy' (ce qui est du ressort de la vie privée), correspond cette fois à la cryptographie dite 'à clé privée', en ce qu'une clé de signature (en général), et parfois de chiffrement, est et doit demeurer privative à son propriétaire et détenteur afin qu'il en soit le seul et unique maître et donc seul et authentique utilisateur.

séclusion et non pas d'apporter une certaine discrétion, car gérer des statistiques anonymes ne doit pas permettre de retrouver un patient, car il ne s'agit pas là de lui apporter des soins ;

- à l'inverse, utiliser des techniques d'anonymisation remplaçant les noms des patients par des numéros plus ou moins aléatoires définitifs et irréversibles, dans un logiciel de gestion de cabinet médical, par exemple, garantirait définitivement et irrémédiablement l'anonymat électronique des patients du cabinet. Mais, on ne pourrait reconnaître les noms ou identités des individus, ainsi répertoriés inutilement. Dans ce cas, il s'agit d'assurer une certaine discrétion dans la gestion des dossiers médicaux ; cette discrétion doit être levée par le médecin, en cas de besoin, ce logiciel de gestion de dossiers médicaux doit précisément de permettre d'accéder aux dossiers médicaux nominatifs, afin de participer au protocole de soins et de favoriser ainsi la continuité des soins.

Sécurité par l'auditabilité

Pour l'auditabilité, il faut obtenir un bon compromis entre perfection et efficacité. L'auditabilité, bien que moins connue et moins pratiquée des quatre propriétés de base de la sécurité est la plus transversale et la plus fondamentale. On recommande, tout particulièrement, son application à tout système d'information d'une sensibilité particulière. Tel est le cas fréquent des systèmes d'information de soins et de santé, lorsqu'il faut fournir à la justice des preuves du bon fonctionnement et de la bonne utilisation du système d'information.

Sans entrer dans des détails élaborés ou fastidieux à ce point de notre réflexion sur la sécurité, il nous faut donner un aperçu clair de ce que l'on entend par *'sécurité par l'auditabilité'* pour ensuite l'aborder utilement sous l'angle de la *'sécurité et intimité des données à caractère personnel'*.

Résumons simplement l'auditabilité[5] d'un système d'information à « sa capacité à fournir en temps et lieu et en forme et fond, les éléments probants nécessaires à la démonstration de son bon fonctionnement et de sa bonne utilisation », c'est-à-dire :

- *capacité :* l'aptitude à être auditable doit être spécifiée, développée et intégrée au système et mise en œuvre de façon opérationnelle ; l'auditabilité se construit au tout premier stade de l'élaboration du système et se poursuit tout au long de son cycle de vie ;
- *en temps et lieu et en forme et fond :* la preuve devra être construite en fonction du contexte d'utilisation du système, de son utilisation probable ou présumée et donc à partir des éléments électroniques probants y afférant, qui devront, par conséquent, être accessibles (cf. *en temps*), accédés (cf. *en lieu*), mais exploitables (cf. *en forme*) et exploités (cf. *en fond*), devant toute instance nécessitant de s'en saisir pour apprécier la réalité de la situation ;
- *les éléments probants nécessaires :* il s'agit des éléments de trace des actions effectuées sur (cf. *'traçabilité'*) ou menées par le système et à la suite de l'imputation de celles-ci à leurs auteurs (cf. *'imputabilité'*), mais aussi, le cas échéant, les éléments *a priori* opposables devant un juge (cf. *'opposabilité'*) et donc les éléments *in fine* acceptés comme irréfutables[6] lors du jugement (cf. *irréfutabilité*) ; on se situe ainsi à la frontière entre la technique et le juridique ;
- *la démonstration de son bon fonctionnement et de sa bonne utilisation :* une démonstration superficielle apporte quelques éléments d'assurance et donc de confiance dans la sécurité du système. Une démonstration plus approfondie utilise une démarche plus formelle : scientifique et/ou juridique : un cheminement rigoureux parcourt les points fondamentaux du droit qui imposent de démontrer l'engagement et donc le respect des responsabilités de chacun des contributeurs au fonctionnement et à l'utilisation du système : commanditaire, spécificateur, concepteur, développeur, intégrateur, testeur, valideur, recetteur, opérateur d'exploitation, utilisateur final,…

L'auditabilité des données à caractère personnel fait apparaître la nécessité de considérer d'autres éléments de confiance indispensables, voire primordiaux pour reconnaître la légitimité et la sincérité des systèmes d'information manipulant des données personnelles.

En voici une liste certes non exhaustive :

---

[5] *La norme européenne expérimentale "ENV13608-1 : Security for healthcare communications – Concepts and terminology" fournit une définition consensuelle de l'auditabilité, suffisamment générique pour être applicable hors du secteur santé.*

[6] *Les juristes se permettront d'utiliser des termes comme irrévocable (action) ou irréfragable (décision).*

- qui définit la politique de collecte, de traitement et d'utilisation des informations nominatives ?
- comment est développée la politique d'occultation, d'anonymisation, de pseudonymisation,… ?
- qui est en charge d'appliquer, de faire appliquer, ou de faire respecter les techniques d'occultation, d'anonymisation, de pseudonymisation ?
- comment est authentifié un individu statistique dans une base de données statistique ?
- qui peut attester de l'identification d'une personne physique parmi une population ?
- qui peut répondre de la fiabilité du processus d'identification des individus d'une population ?
- pourquoi et comment choisir entre confidentialité-discrétion et confidentialité-séclusion ?
- pourquoi, quand et comment lever la discrétion sur des données à caractère personnel ?
- qui doit se porter garant de l'occultation des nomsn de l'anonymisation de données ou de la pseudonymisation d'identifiants ?
- comment garantir la qualité et l'exhaustivité d'informations anonymisées ?
- comment appliquer l'auditabilité à des systèmes d'information non entièrement auditables ?

Les paragraphes qui suivent apportent des éléments de réflexions et de des pistes de solutions, mais ils n'ont pas la prétention de fournir toutes les recommandations et solutions indispensables. Ce serait l'objet d'un tutoriel complet ![7]

Qualité et intégrité des données à caractère personnel

La qualité des processus, en particulier des processus d'identification des personnes physiques, est fondamentale pour la confiance et la fiabilité du système dans son ensemble. Que ce soit de façon purement administrative, lors de l'accueil par l'établissement de soins, lors de l'arrivée pour un séjour d'hospitalisation, que ce soit pour des raisons purement médicales, lors du transfert d'un patient d'un service de soins à un autre, d'une prise en charge avant des soins spécifiques ou pour une opération chirurgicale donnée.

L'intégrité, donc la qualité, des données d'identification est de façon analogue fondamentale et même plus en ce qu'elles permettent au processus d'identification d'être d'un bon niveau de qualité. Une mauvaise identification administrative peut rapidement avoir des conséquences dramatiques, que ce soit en raison de synonymie (enregistrer les informations de l'assuré social dont dépend le patient, au lieu de fournir les informations concernant le patient lui-même, lorsque celui-ci est un ayant droit de l'assuré), ou pour des raisons d'homonymie (enregistrer les informations du patient sur la base d'un séjour d'hospitalisation antérieur d'un autre patient homonyme).

La qualité du processus et l'intégrité des données, surtout en ce qui concerne l'identification des personnes, sont essentiels au bon fonctionnement de l'établissement de soins.

Le débat de mots

L'intégrité des données est souvent confondue avec leur précision et leur exactitude ? Ces notions participent effectivement de l'intégrité des données, mais ne doit-on pas plutôt se poser la question de quand, comment, qui et par qui ces données ont pu être générées et maintenues. Ceci pour apporter toute la confiance nécessaire dans un système d'information sensible, au lieu d'essayer d'atteindre la meilleure précision ou la donnée la plus exacte au risque d'avoir trop de détails sur une donnée déjà devenue obsolète ?

De même, la qualité du système est souvent confondue avec sa fiabilité. Encore faut-il que le système, même s'il est parfaitement fiable, ait été bien conçu c'est-à-dire qu'il prenne en compte les exigences fonctionnelles et les contraintes opératoire auxquelles il doit répondre : un processus très fiable d'identification des personnes ne sera que peu utile si ses conditions de déclenchement ne sont jamais remplies.

---

[7] *Pour plus de détails et une meilleure compréhension, se référer à FD S 97-560, le fascicule de document d'AFNOR qui parcourt exhaustivement, et illustre par divers exemples, toutes les notions et techniques relatives à l'anonymisation.*

Confidentialité(s) des données à caractère personnel

Dans un environnement manipulant des données à caractère personnel, la distinction entre préserver la confidentialité-discrétion et respecter la confidentialité-séclusion devient une question-clé de sécurité.

Voici quelques axes de réflexions et pistes de solutions pour mieux appréhender toute la dimension de cette distinction basique mais essentielle.

Confidentialité-Discrétion (au sens de « secrecy »)

- *Lever la discrétion :* la discrétion est une forme de confidentialité qui pourra toujours être levée si les conditions l'autorise : que ce soit techniquement en faisant appel à des algorithmes cryptographiques réversibles ;
- *Cryptographie symétrique ou asymétrique :* sans vouloir faire un cours de base en cryptographie on peut préciser ici que deux types de fonctions cryptographiques sont couramment utilisés pour mettre en œuvre des outils de confidentialité :
  - les fonctions de chiffrement symétrique ou « à clé secrète ». La clé de chiffrement est maintenue secrète entre l'émetteur du message confidentiel chiffré ; le récepteur de ce message chiffré il pourra le déchiffrer avec la même clé ;
  - les fonctions de chiffrement asymétrique ou « à clé publique ». L'équivalent de la clé secrète précédente est cette fois un bi-clé (couple de clés, l'une privée l'autre publique, corrélées mathématiquement), l'une sert dans l'algorithme de chiffrement et l'autre dans celui de déchiffrement ; selon que l'on se sert de la clé privée pour chiffrer puis de la clé publique pour déchiffrer ou, inversement, de la publique pour chiffrer et de la clé privée pour déchiffrer, on parlera de signature ou, inversement, de chiffrement : dans les deux cas il y a réversibilité en utilisant la clé corrélée à celle qui avait servi à protéger.

Dans les deux cas (symétrique ou asymétrique), il y a, au moins techniquement, une solution pour « déprotéger » ce qui a été précédemment protégé. On pourra toujours prétendre que du point de vue du protocole, de l'organisation au sein du service ou de l'entreprise, ou que, par séparation des pouvoirs, il n'est en pratique pas possible de lever la discrétion. Cependant, déchiffrer des identifiants nominatifs remplacés, pour des raisons de protection des données personnelles, par des données chiffrées restera toujours théoriquement possible, donc de médiocre confiance ;

- *Fragmentation/défragmentation :* la fragmentation-dissémination de fichiers permet de garantir la confidentialité des données éclatées en autant de fragments que nécessaire pour assurer un taux de dispersion optimal ; il faut pour cela associer la fragmentation à la dissémination spatio-temporelle, voire à la redondance par réplication des fragments ou autre technique permettant le recouvrement des données d'origine. L'algorithme de fragmentation a pour but de protéger en confidentialité mais sa spécification permet la reconstitution du fichier en le « défragmentant ».

Ainsi, dans tous ces cas, la confidentialité est maintenue jusqu'à ce que la discrétion soit levée éventuellement mais légitimement, par déchiffrement, par défragmentation ou par toute technique analogue. Ces techniques n'ont pas pour but de faire disparaître les noms ou données à caractère personnel, mais seulement d'en restreindre l'accès en lecture, donc l'usage abusif.

Ces techniques réversibles sont applicables lors d'échanges électroniques, pour l'archivage de fichiers, pour la consultation de bases de données en temps réel ; elles ne se conçoivent que dans l'optique de systèmes d'information mis directement au service des usagers. Les noms, coordonnées, profils, caractéristiques ne figurent dans de tels systèmes que pour la finalité des traitements, comme le rappelle la CNIL. En aucun cas, ces données à caractère personnel, compte tenu des obligations d'information, des droits de rectification, droits à l'oubli, etc., ne doivent servir à d'autres finalités. Mélanger les genres serait préjudiciable pour le respect de la vie privée ; d'où l'utilité de la distinction avec la séclusion.

Confidentialité-Séclusion (au sens de « privacy »)

- *Interdire l'indiscrétion :* les techniques de confidentialité-séclusion doivent permettre de s'opposer à toute forme d'indiscrétion, électronique en l'occurrence, que ce soit directement (fonctions à sens unique) ou indirectement (notions d'inférences) ;

- *Cryptographie à sens unique :* une partie de la cryptographie consiste à mettre en œuvre des algorithmes impliquant une certaine perte d'information de manière à ne pas pouvoir faire l'équivalent d'un déchiffrement comme indiqué dans le chapitre précédent ; en effet, selon la théorie de Shannon, on peut démontrer qu'une certaine perte d'information est nécessaire pour rendre irréversible la transformation de certaine données : classiquement on appelle ce type d'algorithme des fonctions dites « de condensation, hachage ou scellement », à ne pas confondre avec les fonctions de compression/décompression[8].
- *Segmentation :* cloisonnement ou segmentation de sphères de connaissance, annuaires de nommage même anonymisés des individus figurant dans des bases de données sensibles (base de trajectoires de soins, pathologies avérées, prestations sociales,…) ; la segmentation est alors l'homologue de la fragmentation sans aucune technique duale de reconstruction ; dans le respect de la finalité du traitement, la segmentation peut interdire de lever la discrétion ; elle empêche de faire la corrélation entre des faits ou des données qui permettraient de ré-identifier illégitimement les individus statistiques ou qui permettraient de déduire des données sensibles qui ne figurent pas dans la base de données considérée ;
- *Intermédiation :* les australiens l'ont d'abord appelée *'DMA-Data Matching Agencies'*, correspond à une « désegmentation » à la différence près qu'elle doit être spécifiée à l'avance pour être déclarée et légitimée et pour permettre de garder le chemin des segmentations appliquées successivement, de manière à pouvoir reconstruire les corrélations volontairement détruites par la segmentation : sorte de CNIL électronique d'autorisation légale et éthique d'intermédiation entre différents fichiers sensibles.

Dans tous ces exemples, on apportera la preuve, et donc l'auditabilité correspondante, que le nécessaire a été mis en œuvre pour garantir la vie privée des individus, en maintenant la meilleure séclusion possible des données électroniques individuelles, comme l'illustrent les différents concepts et outils pour la séclusion présentés ci après.

Outils pour la séclusion

Anonymat / Anonymie / Anonymité

Il s'agit là de synonymes en français courant (anonymat) ou en vieux françois (anonymie, anonymité) d'un état concernant une donnée ou une information. Il s'agit d'un état de fait et en informatique il faut pouvoir parvenir à cet état de fait : c'est ce que l'on appelle plus généralement l'anonymisation… et autres techniques connexes.

Anonymisation / Pseudonymisation

L'anonymisation et ses diverses variantes, dont l'occultation des noms (éphémère) ou encore la pseudonymisation des identifiants (plus universelle), est une véritable démarche qui inclut une analyse, ou approche, servant un objectif global de respect de la vie privée. Comme indiqué, par la suite, au travers d'un enchaînement de questions pertinentes à se poser (besoins d'anonymat, objectifs d'anonymisation, exigences d'anonymisation), une telle démarche sert à construire la solution d'anonymisation *ad hoc* pour une configuration donnée.

Besoins / Objectifs / Exigences

Les besoins d'anonymisation pourraient presque être multipliés à l'infini, tellement il y a de contextes juridiques, organisationnels ou techniques à prendre en compte : c'est bien souvent la combinaison des réponses à quelques bonnes questions qui permettra alors de cerner le réel besoin : besoin statistique, vision macroscopique ou nécessité microscopique d'entrer dans les détails de données individuelles.

L'objectif, ou finalité du traitement en termes exprimés par la CNIL, contribue à définir quel est le besoin. On peut, selon le FD S 97-560, retenir trois familles d'objectifs : anonymisation réversible (ce qui est un abus de langage, car correspondant plus à de la discrétion provisoire qu'à de la séclusion

[8] *Les fonctions de compression s'appuient, elles aussi, sur la théorie de l'information en exploitant la redondance du langage pour supprimer puis restituer les éléments redondants ou superflus lors de la compression puis décompression.*

définitive), anonymisation irréversible (dans le vrai sens du terme anonymisation) ou anonymisation inversible (sorte d'intermédiaire entre le réversible et l'irréversible).

Les exigences environnementales permettent de compléter l'analyse de besoins. Elles donnent une image fiable de ce que le responsable de la mise en place d'une procédure d'anonymisation est en droit d'attendre. On parlera souvent d'exigence de robustesse, que ce soit robustesse à la réversion non autorisée des données anonymes en donnée (re)nominatives, mais aussi robustesse face à des attaques à base de techniques d'inférence pour déduire, induire des informations confidentielles ou nominatives qui ne figuraient pourtant pas à l'origine dans la base de données manipulée.

Notions de réversibilité : réversible / inversible / irréversible

Trois formes de réversibilités et donc trois familles d'anonymisation ont pu être identifiées :

- *Réversibilité :* possibilité de remonter, depuis les données anonymes ou anonymisées, jusqu'aux données nominatives originelles ; cela peut être envisageable et légitime dans certains cas bien précis, mais mieux vaut éviter d'appeler cela de l'anonymisation pour éviter toute confusion et plutôt parler chiffrement, car telle est la technique qui se cache derrière ;
- *Irréversibilité :* c'est le cas de la réelle anonymisation car décidée et irréversible, comme la séclusion est délibérée et irrémédiable ; une fois remplacé par des identifiants anonymes, les identifiants nominatifs originels ne sont plus recouvrables, de même que remplacées par des pseudonymes les identités ne doivent plus être retrouvables. Cependant, avec les techniques d'attaque par inférence, les pseudonymes s'ils sont trop universellement utilisés risquent de permettre la découverte d'identités mal cachées, comme on l'explicite ci-après ;
- *Inversibilité :* il s'agit là d'un cas mixte entre réversibilité et irréversibilité, c'est-à-dire entre le *'techniquement ou cryptographiquement irréversible'* et le *'organisationnellement et juridiquement réversible'* ; dans ce cas, il faut faire appel à une instance légitime, un tiers de confiance de levée d'anonymat, pour respecter et faire respecter l'intimité électronique de la vie privée des individus concernés. Par analogie, rappelons qu'un représentant de l'Ordre, que ce soit des avocats, des médecins ou des notaires doit être présent lors d'une perquisition, pour faire respect la déontologie et l'éthique de l'Ordre en question, dans l'intérêt du citoyen.

Notions de robustesse : à la réversion / à l'inférence

Nous retiendrons deux types de robustesses.

- *Robustesse à la réversion :* contrairement à la réversibilité la réversion de données n'est pas une capacité souhaitée mais une défaillance face à des risques de perte d'anonymat :
    - ainsi, la réversion peut être directe par des attaques illicites pour retrouver des noms profitant de la faiblesse ou du manque de robustesse de, par exemple, un algorithme cryptographique reconnu pour avoir des failles (fonction à sens unique faillible) ;
    - ou bien, la réversion peut être indirecte si elle profite d'une organisation fragile pour, de façon procédurale, permettre de rétablir la correspondance entre les identifiants anonymisés et les identifiants nominatifs originels et donc violer l'anonymat ;
- *Robustesse à l'inférence :* dérivée de la réversion, l'inférence est une notion qui consiste à dévoiler des informations confidentielles (noms désanonymisés, ou données sensibles) à partir d'autres informations moins confidentielles ou sensibles. Il en existe plusieurs variantes :
    - *L'inférence déductive ou déduction* est la forme la plus connue d'inférence et consiste par la logique souvent booléenne (*valeurs :* oui, non ; *opérateurs :* et, ou) à engendrer des informations simples mais non encore présentes dans la base de données ;
    - *L'inférence inductive* ou *induction* s'apparente souvent à des raisonnements de type loi des grands nombres sans forcément l'appliquer sur de grandes échelles ; cela consiste, par exemple, à induire qui tel patient est très certainement atteint de telle pathologie compte tenu du fait qu'il lui est prescrit tels médicaments comme il est d'usage pour cette pathologie ;

- *L'inférence abductive* ou *abduction*[9] correspond au fait d'extrapoler une situation (une vue partielle de la situation informatique) vers une situation logiquement plus stable, à condition d'ajouter un précepte ou une hypothèse parfois 'tirée par les cheveux' mais finalement fondée : « et s'il avait un cancer, cela expliquerait pourquoi il s'absente parfois d'un Conseil des Ministres pour se rendre du côté de Villejuif… !!! » ;
- *L'inférence adductive* ou *adduction*[10], inverse de l'abduction, permet d'interpréter une situation externe au système informationnel, pour la rapprocher d'une réalité informatique : « il a été vu en consultation à Villejuif et de tous les individus statistiques présents dans la base, c'est le seul : il est donc bien le numéro *XXXX* ».

Notions de chaînage : spatial / temporel et nul / partiel / total

Une autre notion importante, en matière d'anonymisation, est le chaînage d'informations anonymes associées à un même individu dit statistique : épisodes de soins, prestations sociales, prescriptions, symptômes, pathologies, … pour ne prendre que des illustrations issues de la santé ou du social. Ce besoin de chaînage ouvre les risques d'inférence du plus basique (déduction et induction) au plus complexe (abduction et adduction), le chaînage pouvant être spatial / temporel, nul, partiel ou total…

Le chaînage est à la fois un besoin légitime pour mener à bien des travaux et étude, fondés et un risque. Ce risque est encouru dès lors qu'il y a déviance de comportement autour de l'utilisation d'outils informatiques qui permettent de rassembler une grande quantité d'informations privatives sur un individu donné : risque d'atteinte à l'intimité électronique des personnes auxquelles elles appartiennent.

Conclusions

Les propriétés, décrites dans cet article, sont plus ou moins novatrices dans la réflexion de notre communauté ADELI. Par ailleurs, elles sont plus ou moins bien évoquées, et parfois évacuées, faute de temps ou de ressources motivées pour les développer et les remanier.

Ces propriétés ne sont que des associations ou combinaisons, plus ou moins subtiles, d'une kyrielle de propriétés élémentaires à satisfaire puis de techniques de base à assembler et mettre en œuvre. Mais elles répondent à l'un des soucis récents, mais majeurs, engendrés par les nouvelles technologies de l'information et la communication : le respect de la vie privée et la protection de l'individu, dans une dimension électronique qui devient désormais omniprésente.

Avis par courriel : *gilles_trouessin@ernst-young.fr* par téléphone : +33 (0) 6 8255.7406
par la toile : : *www.adeli.com* par télécopie : +33 (0) 1 5847.1033
par courrier : Ernst & Young Audit (France) - Tour Ernst & Young - 92037 PARIS La Défense Cedex

Gilles Trouessin
Vice-président d'ADELI "en charge de l'innovation "
Directeur de Mission au sein de 'Ernst & Young Audit'
Animateur du GE-SSIS et membre des CN IS et CN SSI
Expert au CEN/TC251/WGiii et à l'ISO/TC251/WG4 et 5

AFNOR/GE-SSIS :	Groupe d'Experts *'Sécurité des Systèmes d'Informations de Santé'* d'AFNOR
AFNOR/CN IS :	Commission de Normalisation *'Informations de Santé'* d'AFNOR
AFNOR/CN SSI :	Commission de Normalisation *'Sécurité des Systèmes d'Information'* d'AFNOR
CEN/TC251/WGiii :	Groupe de Travail *'Security, Safety et Quality'* du Comité Technique *'Health Informatics'* du CEN
ISO/TC215/WG4 & 5 :	Groupes de Travail *'Security'* & *'Health cards'* du Comité Technique *'Health Informatics'* à l'ISO

[9] *Par analogie avec un muscle abducteur qui permet à un membre de s'écarter du plan médian du corps.*

[10] *Par analogie avec un muscle adducteur qui rapproche un membre ou un doigt du plan médian du corps, ou de la main.*