

Traitement épidémiologique et démographique du Dossier médical personnel - L'identifiant santé

Catherine QUANTIN¹, Benoît RIANDEY², Olivier COHEN³

¹ Service de Bio-statistique et d'Informatique Médicale, CHU de Dijon,

² Institut National d'Etudes Démographiques (INED),

³ HC-forum.

La loi du 13 août 2004 relative à l'assurance maladie a instauré le Dossier médical personnel (DMP), et prévu sa mise en œuvre pour le 1^{er} juillet 2007. Cette décision, unanimement estimée importante pour la santé publique, a soulevé des débats qui débordent largement du cadre de la coordination des soins : en premier lieu, l'accès des épidémiologistes à ces données, ensuite la question des appariements de données et à cet effet de l'identifiant du DMP ou d'un identifiant unique santé, qui puisse permettre le rapprochement des données de santé. Cette question renvoie également au lien entre données de santé et données démographiques.

1. Partagé ou Personnel ?

La signification même du sigle DMP a donné lieu à controverse : Partagé ou Personnel, ce Dossier Médical ? Si partagé, entre qui ?

D'emblée, on se pose la question du secret médical, le colloque singulier entre le patient et son médecin. Le singulier ne se partage pas sous peine de devenir pluriel !

Des nuances ont dû être apportées à ce principe si fondamental : la loi Huriet a d'abord pris en compte l'intérêt immédiat du patient à la communication entre ses médecins traitants, en premier lieu pour l'établissement d'un diagnostic établi à partir d'analyses ou de l'imagerie médicale. Ensuite, l'intérêt collectif des patients et futurs patients s'appuie sur les progrès de la médecine, notamment dans le domaine de la recherche clinique et épidémiologique. En 1994, la loi Bioéthique a étendu à la recherche médicale ce partage des données sous réserve de l'accord exprès du patient. Or la loi de 2004 énonce une finalité exclusive de coordination des soins et, se référant à la loi Kouchner de 2002 relative aux droits des malades, affirme que ce dossier sera personnel, qu'il sera documenté en accord avec le patient et, sauf cas d'urgence, accessible seulement aux professionnels de santé de son choix. L'éthique et la médecine entretiennent des rapports complexes : s'il ne peut y avoir une bonne médecine sans éthique, certaines dispositions éthiques ne viennent-elles pas contrecarrer l'exercice compétent de la médecine ?

L'information enregistrée dans le DMP ne sera complète que si le patient en décide. Il peut masquer certaines informations à l'égard de médecins postérieurement consultés et masquer le masquage de ces informations (le médecin ignore que le dossier présenté est incomplet) [1]. Voilà une mesure très éthique qui soustrait une partie du sens attribué au mot information, fût-elle médicale. Les expériences de DMP se multiplient dans le monde [2, 3]. Cette option du masquage ne semble pas y être reprise, pour éthiques que soient nos partenaires étrangers. Pour le succès du DMP français, en particulier qu'il reçoive l'assentiment de nos compatriotes, la liberté de s'en soustraire a été souhaitée afin qu'ils puissent y adhérer pleinement. Toutefois, cette disposition n'a pas été retenue pour le DCC, le Dossier Communiquant de Cancérologie, le grand frère du DMP, né du plan national cancer insufflé par le Président de la République. Le cancer est-il trop visible, ou trop grave pour être masqué ? En fait, le succès du DMP auprès des patients est défini comme une affaire de pédagogie. Il s'inscrit dans le temps : au fur et à mesure que l'exhaustivité de la population sera

progressivement atteinte, la complétude des dossiers sera peut-être satisfaite, mais si l'une est vérifiable par simple quantification démographique, le doute subsistera pour l'autre, qu'il soit médecin traitant ou épidémiologiste, voire médecin de santé publique.

2. Et les épidémiologistes ?

Les épidémiologistes ont manifesté leur intérêt pour les données du DMP. Qu'apportent-elles que d'autres sources ne sauraient fournir ? En tout premier lieu, comme les fichiers administratifs, elles présentent l'atout de la gratuité et de l'exhaustivité démographique. En effet, bien que le DMP ne sera alimenté que lors d'un contact avec un professionnel de santé, on peut s'attendre à ce que la population entière soit bientôt couverte, notamment chez les enfants. C'est un avantage considérable par rapport à une enquête par sondage ou, d'un autre point de vue, c'est une source complémentaire ajustée à tout échantillon. Sur le fond, le DMP, partiellement redondant avec le SNIIR-AM [4] pour les consommations remboursables, apportera un éclairage majeur sur le diagnostic des pathologies. Les fichiers de l'assurance maladie n'apportent que des informations bien partielles à ce sujet, sur les affections de longues durées et grâce à la caractérisation médicale de certains médicaments ou au codage. Quant aux enquêtes, elles se limitent généralement au recueil de réponses sur l'état de santé déclaré, tant il est coûteux et difficile d'associer un protocole d'examen clinique.

Les associations de malades ont saisi l'importance du DMP pour la recherche épidémiologique sur les maladies rares, notamment génétiques (la mucoviscidose par exemple) : il permet de constituer un échantillon suffisant de patients pour comparer les traitements et confronter leurs résultats, y compris dans des études internationales comme le propose HC Forum à l'Université de Grenoble.

3. Débat juridique soulevé par les épidémiologistes

Aux questions posées par les statisticiens et les épidémiologistes sur le traitement statistique des données du DMP, les autorités répondaient volontiers fin 2005 - début 2006 que ce traitement est contraire à la loi du 13 août 2004. C'est la raison pour laquelle les associations d'épidémiologistes et de statisticiens, les collèges d'enseignants de Biostatistiques et de Santé Publique ont jugé nécessaire de cosigner un article sur ce sujet, avec l'aide de la CNIL [5]. Dans cet article, nous rappelons que, selon l'article 6 de la loi Informatique et Libertés révisée le 6 août 2004, « *un traitement ultérieur de données à des fins statistiques ou à des fins de recherche scientifique ou historique est considéré comme compatible avec les finalités initiales de la collecte des données* », dès lors qu'il est réalisé dans le cadre des procédures prévues par la loi. La finalité épidémiologique est donc compatible avec une finalité exclusive de coordination des soins dont elle n'est qu'une extension de finalité. Cette loi précise même que, dès lors qu'il ne s'agit pas de prélèvement biologique, l'accès des épidémiologistes aux données médicales n'est plus soumis à accord exprès, mais au simple droit d'opposition du patient dûment informé. Comment une telle nuance sera-t-elle prise en compte dans le décret d'application de la loi du 13 août 2004 ? Le masquage de données sera-t-il applicable aux épidémiologistes ? En particulier lorsqu'ils accéderont aux données de manière anonymisée, notamment grâce à leur technique d'appariement sécurisé ? Il est à parier qu'il reste matière à débat juridique. L'avis de la CNIL sera à juste titre déterminant, et on sait à quel point elle est attentive aux missions de santé publique dès lors que les garanties d'anonymisation sont apportées. D'ailleurs comment pourrait-on évaluer la pertinence des choix inscrits dans la loi du 13 août 2004, en particulier de son efficacité pédagogique, si aucun traitement statistique ne vient mesurer la fréquence des dossiers qui connaissent un masquage partiel, la fréquence du masquage par pathologie, le caractère systématique ou aléatoire du masquage ?

Il est aujourd'hui reconnu que l'accès au DMP sera ouvert aux épidémiologistes, mais pas au début (cf. colloque national DMP du 4 décembre 2006). Cette réserve, pour erronée qu'elle soit juridiquement, n'est guère préoccupante car les épidémiologistes n'auront pas d'intérêt propre dans le DMP avant qu'il n'ait atteint une certaine maturité quantitative et qualitative et que, de plus, on viendra les chercher pour évaluer cette maturité qualitative. Ne vaut-il pas mieux se faire prier que d'être quémandeur ?

4. Vous avez dit appariement ?

L'analyse statistique des données du DMP ne saurait constituer une fin en elle-même, si ce n'est pour évaluer le dispositif du DMP. Son intérêt réside avant tout dans le rapprochement d'autres sources, notamment pour les chercheurs d'autres disciplines comme les économistes et les démographes.

L'évaluation de la politique de santé publique comprend nécessairement celle de l'allocation des ressources financières aux différentes pathologies. Les diagnostics du DMP sont donc fort utiles aux données d'assurance maladie, qu'il s'agisse de tirer du DMP un échantillon de patients atteints d'une affection donnée pour l'apparier aux dépenses occasionnées ou au contraire d'enrichir avec les diagnostics du DMP un échantillon tel que l'EPIBAM¹ [3]. En définitive, il s'agit presque toujours d'enrichir un échantillon avec une extraction de variables issues d'un fichier exhaustif.

Pour les démographes, le DMP contribuerait d'abord à l'analyse de la mortalité par cause grâce à une observation précise, longitudinale et donc datée des pathologies, causes premières sinon absolues des décès. Bien sûr, ce lien entre pathologie et décès sera d'abord le fait de services hospitaliers qui s'interrogent sur la survie de cohortes de leurs patients. L'étude d'Isabelle Fournel [6] pour l'Institut Gustave Roussy, présentée à ce même colloque en est un exemple caractéristique. Elle nous rappelle comme les informations d'état civil constituant le NIR sont utiles pour un tel rapprochement : la date de naissance complète, le département et la commune de naissance. Par rapport aux fichiers d'assurance maladie, le DMP présente l'avantage d'éviter de choisir entre profondeur temporelle, l'atout futur de l'EPIBAM et l'exhaustivité démographique du SNIIR-AM limitée à trois ans de consommation médicale.

C'est un des enjeux du prochain décret relatif au DMP : la durée d'archivage des données, notamment pour les maladies génétiques dont le diagnostic intéresse la (les) génération(s) future(s). Des voix se sont, en effet, élevées pour étendre les durées de conservation des données prévues par le premier projet de décret. Le droit à l'oubli, reconnu par la CNIL comme un principe fondamental en même temps qu'une précaution protectrice, affaiblit la discipline démographique et désarme l'épidémiologie analytique.

Sans doute une réflexion plus approfondie sur les champs connexes à l'épidémiologie et à la démographie ajouterait de nombreuses utilisations potentielles du DMP. Par exemple, pour les démographes, dans le domaine de la fécondité assistée ou des sous-populations handicapées. Pour s'en tenir à la démographie générale, par son exhaustivité démographique, le DMP fournira un des indices de présence sur le territoire ; ce serait particulièrement intéressant à l'égard de quasi non consommateurs de soins lorsque seraient enregistrés des actes médicaux gratuits ou non remboursés. Dans l'idée d'utiliser le RNIAM² comme un pseudo registre démographique ou le SNIIR-AM comme un indicateur de présence sur le territoire [7], on bute sur la sous-observation des jeunes adultes en bonne santé trop facilement supposés partis à l'étranger. Toute indication de présence peut corriger des biais d'observation comme par

¹ Échantillon permanent inter-régimes des bénéficiaires de l'assurance maladie.

² Répertoire national inter-régime de l'assurance maladie, géré par la CNAV en coordination avec le répertoire de l'état civil de l'INSEE, le RNIPP.

exemple, une visite médicale gratuite pour un contrôle dentaire sans remboursement. L'identifiant crypté assorti d'une date constitue un indice de présence précieux, au même titre qu'une date d'inscription en formation³ ou à un fichier de salariés comme les DADS⁴.

Les demandes d'appariement du DMP relèveront d'abord bien sûr des épidémiologistes. Comme le signalait Marcel Goldberg [8], leur travail présente d'abord une grande exigence de validation par comparaison de sources. Ce serait la première utilisation épidémiologique d'un DMP encore en défaut d'exhaustivité. Cette confrontation des sources demande appariement et donc identification.

5. Quel identifiant pour le DMP ?

Cette question a sans doute créé le plus chaud débat autour du DMP. Ce n'était pas faute de précaution puisque, dans un but préventif, le parlement avait prévu que le DMP ne pourrait être lié à aucun autre fichier notamment d'assurance maladie. Son identifiant, l'Identifiant National de Santé ou INS (initialement dénommé NIS), ne serait qu'un numéro d'ordre établi après contrôle de la primauté de la demande à l'aide du RNIAM, mais sans conservation de la table de correspondance entre NIR et INS. Cette politique voulue prudente a au contraire fait réagir les institutions les plus pondérées, jusqu'aux académies des Sciences [9] et de Médecine [10], sans oublier les six associations d'épidémiologistes ou de statisticiens qui ont uni leur appel dans un article de la RESP [5] : un DMP nominativement identifié aurait soulevé des vagues ; le projet de DMP à l'identifiant muet en a soulevé symétriquement.

Le changement de direction à la tête du GIP DMP a abouti à un revirement complet sur ce point, mais dans un premier temps non pacificateur. L'analyse partait d'un constat réaliste : ajouter sur la carte Vitale et dans les vastes fichiers hospitaliers si mal protégés un INS à côté du NIR ne créerait qu'une illusion de sécurité beaucoup plus dangereuse que son absence ; pensons à une rambarde non fixée sur un chemin de crête. Autant prendre le NIR comme identifiant du DMP. Économistes de la santé et épidémiologistes verraient leur demande satisfaite, et pour ajouter à la coordination des soins, instituons le NIR comme identifiant national santé commun à tous les fichiers médicaux. La protection du secret médical devait provenir d'une sécurisation des accès et d'une pénalisation lourde des infractions commises. La CNIL consultée sur cette proposition audacieuse l'a réfutée comme contraire au texte de la loi du 13 août 2004. En décembre 2006, le gouvernement a donc déposé un amendement à la loi de santé publique instituant un INS et chargeant la CNIL de proposer sa nature.

Cette position engagée du ministère a présenté deux avantages : celle d'amener la CNIL à examiner les fichiers hospitaliers et à constater leur niveau de protection insuffisante, faute d'avoir attribué à ses agents la carte de professionnel santé diffusée par le GIP CPS aux professionnels de santé du secteur privé⁵. L'autre avantage de cette position extrême a été de placer les épidémiologistes revendiqueurs en position implicite de médiateurs.

6. Pour un identifiant sécurisé, hachons notre NIR

En tant qu'identifiant du DMP, le NIR permettrait les appariements utiles du DMP sans qu'il corresponde au choix des académies et associations mobilisées. Les épidémiologistes ont abordé ce débat par la porte du DMP, c'est à dire sa connexion aux bases existantes, sans

³ Grâce à l'association d'un NIR haché et de l'identifiant INE de l'Éducation nationale. L'idée originale selon laquelle le NIR haché pourrait servir à dédoubler l'INE trouverait là une application de premier ordre.

⁴ *Déclarations annuelles de données sociales*, antérieurement *Déclarations annuelles de salaires*, faites par les employeurs de l'année auprès des institutions sociales.

⁵ Le 15 mai 2007 est pris un décret (Art L110-4) créant sous trois ans obligation de la carte CPS à tous les professionnels y compris hospitaliers accédant aux données médicales.

engager le débat sur les déficiences actuelles du système. C'est à l'honneur du ministère de la santé d'avoir lancé le pavé dans la mare en élargissant la question des épidémiologistes à toutes les difficultés de communication entre les réseaux de santé, et d'abord entre hôpitaux publics.

Les épidémiologistes ont seulement réclamé que le lien entre le DMP et le NIR ne soit pas coupé par l'instauration d'un numéro d'ordre dont l'unicité serait validée par le RNIAM sans conservation de la table de correspondance entre NIR et numéro d'ordre. La proposition était seulement de crypter et compresser de façon irréversible le NIR (on dit hacher) pour éviter toute réidentification [11, 12], tout en permettant les appariements avec les fichiers du SNIIR-AM, du PMSI et de l'EPIBAM qui utilisent la même famille d'identifiants FOIN [13]. La question des grandes cohortes épidémiologiques était posée sans que soit énoncée une solution globale. Le ministère l'a fait. C'est donc le moment de développer notre analyse pour favoriser une solution complète au problème posé.

7. Débat sur l'identifiant et arbitrage de la CNIL

Le débat, important sur le fond, a soulevé des discussions animées : les associations de Droits de l'Homme ou de patients ont appelé à pétitionner contre le choix du NIR comme identifiant santé, faisant valoir que le traitement administratif des remboursements était d'une toute autre nature que le traitement médical. Ils souhaitaient que, dans une finalité pédagogique, un nouvel INS soit ajouté sur la carte Vitale 2. Parallèlement, ils reconnaissaient que le NIR haché satisfaisait à leur attente. D'un autre côté, le ministère souhaitait une solution efficace qui ne se cache pas derrière la fausse sécurité d'un INS voisinant le NIR sur toutes les cartes vitales ni dans les fichiers hospitaliers mal protégés comptant des millions de patients. Il reconnaissait que le NIR haché satisfaisait cette exigence.

Tout le monde était-il d'accord ? Pas du tout, car le GIP-DMP souhaitait que l'on garde le NIR le plus possible, et que, si on devait le hacher, on ne rajoute pas l'INS ou NIR haché dans la carte vitale, pour des raisons de sécurité de l'INS et de coût de déploiement (modification de la carte vitale et des systèmes d'information des professionnels de santé). Il trouvait que cette solution très compliquée, source d'erreur, consistait à déposer un leurre sur la carte Vitale avant de faire à nouveau un hachage pour sécuriser le NIR lui-même. Les associations s'y opposaient par souci pédagogique et souhaitaient que chaque citoyen puisse visualiser que l'on n'utilise pas le NIR mais un nouvel identifiant. Il a donc été demandé par les associations qu'un INS côtoie le NIR sur la carte vitale avant d'être haché pour des usages sécurisés.

Dans son avis du 20 février 2007, la CNIL a tranché : rappelant que le NIR n'est pas un identifiant comme les autres et que les fichiers de santé ne sont pas des fichiers comme les autres, la CNIL a déclaré nécessaire que l'INS ne soit pas le NIR, mais que ce serait le NIR haché.

Les associations ont solennellement appelé le gouvernement à ne pas s'écarter de l'avis de la CNIL. En définitive, la décision est prise de suivre l'avis de la CNIL. La CNAV sera chargée d'assurer la transformation du NIR par algorithme de hachage. Au cours des deux prochaines années, l'INS (Identifiant National de Santé) ou NIR haché sera inclus dans la carte vitale 2. Les associations ont donc eu satisfaction sur ce point. Cet INS serait haché à nouveau à l'arrivée sur le portail du DMP. De l'eau coulera sous les ponts avant que ce projet ne soit consolidé. L'avenir peut donc encore contredire ces propos tenus fin mai 2007 au colloque de la CUDEP.

Les épidémiologistes s'attellent à approfondir les contours de cette solution. Par delà son aspect narratif, cette communication essaie d'y contribuer. Les épidémiologistes ont demandé un identifiant spécifique FOIN2 pour le DMP. Ce message important est-il passé ?

En premier lieu, pour assurer la compatibilité avec les données du Système National d'Information de l'Assurance Maladie (SNIIR-AM) l'identifiant du DMP devrait relever de la

branche FOIN2 de cette famille d'identifiants. Ensuite, nous constaterons qu'il ne s'agit pas seulement d'un identifiant santé mais d'un système « FOIN 2 » d'identifiants santé.

Les épidémiologistes préconisent également le rajout du nom, prénom, date de naissance (hachés), afin d'une part d'être compatible avec l'identifiant de la carte européenne de santé, et d'assurer l'interopérabilité avec l'identifiant européen lorsqu'il y en aura un, mais aussi afin de permettre la corrélation des données du DMP avec les données des professionnels de santé notamment des établissements hospitaliers : en effet, ceux-ci risquent d'avoir du mal à faire le lien entre leurs identifiants (qui ne comportent pas actuellement le NIR mais le numéro de sécurité sociale de l'assuré) et l'INS.

8. De FOIN à FOIN2

Rappelons que l'identifiant de la carte Vitale 1, implémenté dans le PMSI et le SNIIR-AM est constitué du NIR de l'ouvrant droit complété de la date de naissance complète du bénéficiaire et de son sexe [13]. Le tout est haché en un bloc par la fonction FOIN. Il en ressort une chaîne de 40 caractères susceptible des seuls traitements informatiques.

Cette *fonction d'occultation des informations nominatives* est à la base de systèmes d'information considérables de l'assurance maladie, mais il présente deux faiblesses qu'il serait temps de corriger. Une malencontreuse décision hiérarchique a substitué le code sexe (1 ou 2) à un indicateur de gémeité valant 1 pour une grossesse unique et de 1 à 5 pour des quintuplés. La fonction FOIN ne sait pas reconnaître des jumeaux, soit un cinquantième de la population environ. En contrepartie, elle a permis de distinguer des conjoints dont l'un serait ayant droit de l'autre, par ailleurs nés le même jour de la même année. Le gain informatif est fort maigre pour une correction qui semblait se conformer à l'évidence ! En second lieu, la variable sexe est si souvent erronée que son introduction produit de nombreux NIR hachés faux, donc autant de doublons résistant à l'appariement dans la base.

La diffusion de la carte Vitale2 est commencée, dans un premier temps sans l'INS, mais avec un identifiant fondé sur le NIR du bénéficiaire.

9. Plaidoyer pour un ICI santé

Comme nous l'avons vu dans la partie « Débat sur l'identifiant et arbitrage de la CNIL », la sécurité des fichiers de santé n'est guère assurée par l'introduction d'un identifiant santé unique dans la carte vitale. Certes, il est constitué d'une chaîne de 40 caractères immémorable, ce qui empêchera quiconque de le reconstituer. C'était bien un avantage du NIR que son caractère significatif permette à chacun de le retenir : le patient est même capable de l'écrire sans se tromper. Il n'en sera pas de même du NIR haché, au point qu'on pourrait imaginer de lui donner un « correspondant » plus civilisé.

En fait, la coordination entre fichiers ne doit être effective qu'à l'issue d'un feu vert reconnu, comme pour tout nouveau type d'appariement autorisé par la CNIL. C'est bien ce que mettent en place les autorités belges en distinguant l'identifiant des fichiers médicaux de recherche de ceux de soins ... Un identifiant par application et un système protégé de clés de chiffrement qui permette à ces applications de communiquer. C'est le système général du DMA, le Data Matching Agency australien, ici comme en Belgique appliqué au seul domaine de la santé. En français, nous l'intitulons ICI, Instance de Coordination des Identifiants.

10. ICI santé et CNIL

Actuellement la CNIL décide de tout appariement. La montée en puissance des appariements sécurisés, avec les précautions à prendre sur les risques des fichiers indirectement

nominatifs fait penser qu'une instance sous le contrôle étroit de la CNIL ne serait aucunement superflue.

L'expérience du ministère des finances relève de ce domaine. Actuellement, l'application mise en œuvre par l'Éducation nationale réalise ses appariements de fichiers par voie strictement interne. Même le dédoublement de son identifiant INE peut se gérer en interne. Mais pourquoi ne voudrait-elle jamais accéder à des fichiers externes ou en transmettre pour appariements ? Grâce à la mise en place d'une ICI, la CNIL autoriserait le traitement et l'instance gérerait la coordination des clés de chiffrement.

La CNIL est très motrice dans ce domaine. Au terme du cours d'appariements sécurisés de la SFdS, le 16 novembre dernier, rassemblant 60 statisticiens d'institutions les plus diverses, la directrice du service juridique de la CNIL concluait : « *Statisticiens ayez de l'audace* » ! On ne peut qu'appuyer cette incitation à utiliser pleinement les techniques modernes productives et sécurisées. La conséquence en serait qu'il faut s'organiser.

Plaçons-nous dans une perspective plus longue : c'est l'aboutissement, la maturité de la politique de la CNIL prônant la mise en place d'identifiants spécialisés. Quand dans les années 80, elle a défini cette politique, sa mise en œuvre s'est faite sans que les ministères disposent de dispositif leur permettant de valider leur identifiant et de communiquer entre secteurs, lorsque ce transfert s'avérait utile et autorisé. L'Australie a créé son DMA à cet effet. Sans doute ce dispositif n'était-il pas acceptable tel quel en France, comme les appariements moins sophistiqués des statisticiens nordiques. La libéralisation des fonctions de cryptage, le fait qu'un hachage sans retour ne permettait pas de voiler temporairement des informations dangereuses pour la sécurité nationale, le progrès méthodologique ont ouvert une voie apportant toute garantie éthique. Il n'y a donc plus de raison de rechigner contre la sectorisation des identifiants si le dispositif précédent vient à maturité.

Il est remarquable que la Ligue des droits de l'Homme ait promu cette solution lors des débats relatifs à la transcription de la directive européenne de 1995. En vain, malheureusement. À cette époque, la Société française de statistique s'intéressait moins aux identifiants qu'aux sondages et à la reconnaissance de la finalité statistique. Un échange plus précoce aurait peut être permis le 6 août 2004 de faire inscrire l'institution d'une ICI dans la révision de la loi Informatique et Libertés.

11. Et les démographes dans tout cela ?

Leur premier devoir, comme aux statisticiens, est d'apprendre ces techniques. Viendront ensuite une multitude d'applications ouvertes aux audacieux. Prenons l'exemple des pacs et dépacés : Destinée à évaluer la loi de 1999 légalisant les unions homosexuelles, la statistique du pacs, sept ans plus tard, n'avait encore fourni aucune indication de la part des unions homo ou hétérosexuelles ainsi légalisées. Cette lacune justifie pleinement l'appréciation de Patrick Festy « *Pacs, l'impossible bilan* » [14] ! Pourtant, forts du serment d'Hippocrate, les épidémiologistes avaient précédé leurs collègues statisticiens sur ce sujet en mettant en œuvre à l'INSERM dès 1988 puis à l'InVS en 2001 [15] une statistique sans double compte des séropositifs VIH sida précisant le mode de contamination (hétéro, homo, toxicomanie). Pourquoi n'a-t-il pas encore été débattue identiquement l'opportunité d'un registre anonyme sécurisé de pacs fondé sur le hachage de l'identification des personnes et mentionnant la composition sexuée du couple ? Par la même voie, les ruptures de pacs seraient immédiatement appariées avec le pacs rompu. Ainsi obtiendrait-on directement une statistique plus moderne que celle qui ne relie pas encore aussi simplement mariage et divorce.

La CNIL avait été la première institution à avoir recommandé l'usage des techniques épidémiologiques de hachage pour la statistique publique. C'était pour la création de l'Observatoire du RMI à Paris réalisé par le CREDOC [16]. Puis, à la suite d'un premier cours

de la Société française de statistique, le Ministère de l'éducation nationale a appliqué ces techniques aux statistiques d'étudiants [17]. À nouveau la CNIL a repris son message à l'occasion de ses *Dix recommandations* pour la mesure de la diversité [18]. Elle recommande à la fois les techniques de cryptage et le recours à un tiers de confiance. En prônant le hachage et la mise en place d'une/des instances d'ICI, nous nous situons bien dans le sillage de la CNIL. C'est exactement de cas de l'équipe ELFE dirigée par Henri Leridon [19]. Sa grande cohorte d'enfants, bientôt lancée en 2009 par une enquête de périnatalité, puis insérée dans l'Échantillon démographique permanent élargi de l'INSEE, cumulera une grande diversité de données issues de sources nombreuses. S'interdisant de créer une mégabase de données ELFE, l'équipe prévoit exactement la mise en place d'une ICI qui gèrera la mise à disposition des données requises et réalisera les appariements minimaux ad nominem ad hoc. Le potentiel de ces méthodes est donc très prometteur dans le domaine de la démographie.

Citons aussi la génétique des populations dont les noms du docteur Jean Sutter et d'Albert Jacquard témoignent de l'affinité avec la démographie, sans oublier l'anthropologie dont les généalogies ont été si utiles aux travaux des généticiens. C'est bien l'objet de l'identifiant à composante familiale que nous avons proposé (Catherine Quantin, Olivier Cohen) dans le cadre d'un brevet international, qui associe à l'identifiant de l'enfant celui de chacun de ces parents⁶. Son intérêt est évident pour les recherches sur les maladies génétiques rares, recherches nécessairement internationales. D'où l'idée de cet identifiant familial européen que nous avons proposé lors du colloque de l'association européenne d'informatique médicale [20]. Il permet aussi de renouer avec la dimension généalogique souvent oubliée par la démographie. En témoigne l'expérience engagée en commun sur un sujet aussi sensible que l'enfance en danger par l'ONED, HC Forum et le CHU de Dijon : cette application permet, de façon anonyme et avec l'accord de la CNIL, de reconnaître des enfants en danger d'une même famille et de reconstituer leurs parcours de prise en charge. Un objectif qu'on ne se serait pas permis d'envisager dans un passé proche en l'absence de ces techniques.

Conclusion

Les médecins et épidémiologistes ont apporté des solutions efficaces et sûres aux besoins de rapprochement de données, y compris dans la dimension politique et juridique du débat. Les démographes en seront également bénéficiaires. À leur tour de s'y investir.

BIBLIOGRAPHIE

- [1] FAGNIEZ P-L., (2007), « le masquage d'information par le patient dans son DMP » - Rapport au ministre de la santé et des solidarités.
- [2] BOURQUARD K, (2007), « DMP : Situation Internationale » *Pratiques et organisations des soins*, n° 1, CNAMTS.
- [3] QUANTIN C., COHEN O, RIANDEY B., ALLAERT F.A. (2007), « Unique Patient Identifier : a world wide review » *International Journal of Medical Informatics*.
- [4] LENORMAND F., (2005), « Le système d'information de l'assurance maladie, le SNIIR-AM et les échantillons de bénéficiaires ». *Courrier des Statistiques*, 2005, N°113-114 : 33-51

⁶ Cet identifiant est alors haché par morceaux par la fonction Anonymat développée au CHU de Dijon [11]. Cette technique probabiliste sécurisée protège d'erreurs d'identification.

et Journal de la SFdS vol 146, n°3, chaînage entre des données hospitalières et les données de mortalité nationales anonymisées.

- [5] QUANTIN C., GUINOT C., TURSZ A., SALOMEZ J.-L., ROGIER C., SALAMON R. (2006), « Le traitement épidémiologique du Dossier Médical Personnel au service des malades » ? *Rev Epidemiol Sante Publique*, 54.
- [6] FOURNEL I., (à paraître dans la *Revue d'épidémiologie et de santé publique*), « Détermination du statut vital par chaînage entre des données hospitalières et les données de mortalité nationales anonymisées ».
- [7] RIANDEY B., (2007), « Le nouveau système d'information de l'assurance maladie en France permettrait-il une estimation des flux d'émigration ? » in « Les migrations internationales. Observation, analyse et perspectives », Colloque de l'AIDELF à Budapest, septembre 2004.
- [8] GOLDBERG M., (2006), « Les bases de données d'origine administrative peuvent-elles être utiles pour l'épidémiologie ? », *Revue d'épidémiologie et de santé publique*, vol. 54, n° 4, p. 297-303.
- [9] ACADÉMIE DES SCIENCES, (2005), Rapport « Science et Technologie ». Comité RST « Épidémiologie : conditions de son développement, et rôle des mathématiques ». Animateur : Alain-Jacques Valleron. 11 janvier 2005.
- [10] ARDAILLOU R, HENRION R., (2006), Communiqué sur l'utilisation en épidémiologie des bases de données recueillies en population générale, 14 février.
- [11] QUANTIN C, BOUZELAT H, ALLAERT FA, BENHAMICHE AM, FAIVRE J, DUSSEYRE L. (1998), Automatic record hash coding and linkage for epidemiological follow-up data confidentiality. *Methods of Information in Medicine*. 37 :271-277.
- [12] QUANTIN C, GOUYON B., ALLAERT FA, DUSSEYRE L. COHEN O., (2005), « Méthodologie pour le chaînage de données sensibles tout en respectant l'anonymat : application au suivi des informations médicales ». *Courrier des statistiques*, 2005 ; 113-114 :15-26 et *Journal de la SFdS*, vol 146, n°3.
- [13] TROUËSSIN G., ALLAERT F.A. (1997), FOIN : a nominative information occultation function », *MIE*, 3, 196-200.
- [14] FESTY P. (2001), « Pacs, l'impossible bilan », *Populations et sociétés* n° 369.
- [15] CNIL (2001), « 21^{ème} rapport d'activité 2000. Délibération n° 00-031 du 25 mai 2000 ». La Documentation française. (Déclaration obligatoire du sida auprès de l'InVS).
- [16] ALDEGHI I. et OLM Ch. (2004), « L'observatoire des entrées et sorties du RMI à Paris » in Ardilly P. (dir) « Échantillonnage et méthodes d'enquêtes », Actes du 3^{ème} colloque francophone sur les sondages de la Société Française de Statistique, Dunod, Paris.
- [17] GOY A., « L'appariement sécurisé des fichiers d'étudiants grâce au hachage des identifiants ». CNIL (2007), Dossier spécial « panels et appariements sécurisés », *Courrier des statistiques* n°113-114, juin 2005 et *Journal de la Société française de statistique*, vol 146 n°3, 2005.
- [18] CNIL, « Mesure de la diversité et protection des données personnels. Les dix recommandations de la CNIL ». Lettre INFOCNIL, mai 2007.
- [19] CNIS (2007), « Note de présentation du projet Elfe », Réunion de l'inter-formation Statistiques Sociales du 6 juin 2007 sur « Elfe ».
- [20] QUANTIN C. ALLAERT FA, GOUYON B, COHEN O., « Proposal for the creation of a European healthcare identifier », MIE Genève 2005. *Stud Health Technol Inform*. 2005 ; 116 : 4-54.