



L'identifiant et le chaînage de l'information

**Comment choisir
un identifiant
pour l'utilisateur
du DMP qui permette
le chaînage individuel
des données en vue
d'un traitement
épidémiologique
au service
des patients ?**

Catherine Quantin
Professeur, CHU de Dijon,
service de Biostatistique
et Informatique médicale,
département d'information médicale

Dans le cadre des réformes de l'hôpital 2007, le dossier médical personnel (DMP) représente une avancée très importante pour la coordination des soins. La France est l'un des premiers pays au monde à envisager un tel dispositif pour l'ensemble des bénéficiaires d'une assurance maladie. Sur le plan de la santé publique, les enjeux dépassent l'objectif initialement fixé, dans la mesure où le recueil de données de santé à l'échelle d'une population permet d'envisager des études épidémiologiques de grande ampleur avec des retombées pour la prise en charge des malades, que ce soit par sa contribution à la recherche médicale ou à la planification sanitaire, notamment pour les schémas régionaux d'organisation sanitaires (Sros). Toutefois, tous ces enjeux, que ce soit pour la coordination des soins ou pour les études épidémiologiques, supposent d'avoir mis en place un identifiant du patient qui soit suffisamment fiable et qui permette le respect des législations européenne et française sur le traitement des données à caractère personnel.

Le mode de génération du NIS dans le DMP de préfiguration ?

En septembre 2005, le comité d'orientation du GIP-DMP a défini les principes fondateurs du numéro d'identification santé (NIS) du dossier médical personnel : unique, non signifiant, public, pérenne, irréversible.

Pour respecter la condition d'unicité, le groupe de travail Identification du patient a suggéré au comité d'orientation d'associer l'identifiant de la personne au NIR, numéro d'inscription au Répertoire national d'iden-

tification des personnes physiques (RNIPP), géré par l'Insee. La solution proposée par ce groupe de travail s'appuie sur la génération d'un numéro d'identifiant santé pour chaque patient, géré à sa demande par un organisme indépendant appelé « tiers de confiance ». Le rôle du tiers de confiance, qui ne doit pas être un hébergeur, est de garantir que l'identifiant est sans doublon.

Pour préserver l'anonymat de cet identifiant, ce tiers de confiance crée le NIS à partir du schéma suivant : le patient (le bénéficiaire) choisit un hébergeur agréé, en avertit le guichet du DMP par l'envoi d'un bordereau de la demande de NIS, ainsi que son hébergeur en lui adressant un formulaire d'adhésion. Le guichet vérifie auprès du Répertoire national inter-régimes des bénéficiaires de l'assurance maladie (RNIAM) l'affiliation du demandeur et la primeur de sa demande, et transmet, le numéro de bordereau au tiers de confiance. Le tiers de confiance crée le NIS à partir du numéro de bordereau et le transmet, avec les autres traits d'identification du patient tels que son nom, son prénom, sa date de naissance, à l'hébergeur. L'hébergeur envoie alors au patient l'adresse qualité santé (AQS). Cette AQS composée du NIS et de l'identifiant de l'hébergeur, est communiquée, par le patient, aux professionnels de santé qu'il consulte.

L'utilisation du numéro de bordereau et non pas du NIR pour la génération du NIS est censée garantir, d'une part, son caractère non signifiant et, d'autre part, l'irréversibilité mathématique de la procédure de création du NIS. En effet, à partir du NIS, aucune opération mathématique ne permet de remonter directement au NIR.

Les critiques à apporter à ce mode de génération

Les principales critiques [1] concernent le fait que la solution proposée conduit à déconnecter institutionnellement le NIS, proposé en tant qu'identifiant central du système de santé, par rapport au NIR, dans le souci de ne pas construire une base de population contenant la table de correspondance entre le NIR et le NIS.

Toutefois, le fait que la relation entre le NIR et le NIS soit définitivement coupée rend impossible le rapprochement des données du DMP avec d'autres données concernant le patient à des fins de recherche. Le rapprochement des données du DMP avec d'autres sources est nécessaire non seulement pour compléter mutuellement leurs données mais également pour assurer leur validation multi-sources, conformément aux recommandations de l'Académie nationale de médecine (14 février 2006) et au rapport de l'Académie des sciences relatif à l'épidémiologie [2]. Si l'on sait que la recherche sur les données, du DMP ne sera pas possible dans l'immédiat, du fait du temps nécessaire à la constitution de bases de données suffisamment structurées, exhaustives et de qualité validée (comme par exemple, pour le PMSI), on ne peut pas attendre que ces données soient utilisables pour constituer alors un identifiant acceptable. En effet, on sait bien quelles sont les complications techniques et économiques générées par un changement d'identifiant pour de nombreuses applications nationales. En outre, le dossier communicant de cancérologie (DCC) doit se mettre en place avec le même calendrier que le DMP et avec le même identifiant. L'objectif du DCC étant clairement de fournir aux professionnels de santé un outil pour les soins et la recherche, il devient indispensable que l'identifiant imposé pour le DCC et le DMP permette de répondre à cet objectif.

Les solutions préconisées pour le DMP

Il serait tout à fait possible de préserver la confidentialité due au patient et souhaitée par les associations de patients en mettant en place des systèmes d'anonymisation [3] tels que celui adopté par l'Institut de veille sanitaire, sur recommandation de la Cnil, dans le cadre de la procédure de suivi des 30 maladies à déclaration obligatoire (dont le sida). Contrairement aux méthodes de cryptage qui doivent pouvoir être réversibles de façon à ce que le destinataire légitime puisse déchiffrer le message, les méthodes

cryptographiques de hachage à sens unique sont irréversibles. Le résultat du hachage est un code strictement anonyme (ne permettant pas de revenir à l'identité du patient), mais toujours le même pour un individu donné de façon à pouvoir rapprocher les données d'un même patient. Les applications dans le domaine médical sont nombreuses, avec la constitution de bases de données nationales (telle celle relative au suivi des personnes séropositives qui fréquentent les CISIH, qui concerne actuellement environ 100 000 patients et qui est un bon exemple de ce qui peut être mis au service de la recherche épidémiologique, avec le plein accord des patients), régionales ou interrégionales dans divers domaines (cancer, périnatalité, maladies génétiques). Ce système a également permis le chaînage des données du Programme de médicalisation des systèmes d'information (PMSI) et du SNIIR-AM, système d'information de l'assurance maladie.

Dans le cas du DMP, la situation n'est en fait pas plus complexe car il s'agit de la même manière de répondre à la fois :

- aux exigences de la Cnil, des associations de patients et des professionnels de santé en matière de confidentialité des informations personnelles contenues dans le DMP, et dans le respect de la loi du 13 août 2004 concernant l'utilisation de ces données,
- aux garanties apportées aux professionnels de santé d'une anonymisation des données du DMP à leur égard¹,
- au besoin en matière de santé publique ou à titre individuel de disposer de ces données, notamment lorsque le patient a donné son consentement exprès.

L'idéal serait d'utiliser l'identifiant du SNIIR-AM, compatible avec l'identifiant du système FOIN pour le PMSI, qui permettrait d'assurer le rapprochement avec les principales bases de données nationales. Le lien entre le NIS et le NIR serait donc rompu, dans le sens où il ne serait pas possible, à partir du NIS, de reconstruire directement le NIR. Toutefois, il serait possible d'apparier des données du SNIIR-AM avec des données du DMP, à condition d'avoir l'autorisation de la Cnil. On pourrait, par exemple, utiliser une clé de hachage supplémentaire si l'on souhaitait éviter de stocker les données du DMP avec

le même numéro d'identifiant que celles du SNIIR-AM. Cela reviendrait à mettre en place des identifiants distincts par finalité (administration de la santé, diagnostics et soins, recherche épidémiologique) dérivés par hachage (ou algorithme de cryptage irréversible) à partir d'un numéro de Sécurité sociale unique, comme ce qui est proposé en Belgique [4] ou en Nouvelle-Zélande.

Le rapprochement pourrait alors être fait, de manière sécurisée, avec les données nationales du PMSI. En effet, les données du PMSI, lors de la transmission par les établissements hospitaliers à l'agence régionale de l'hospitalisation, puis au ministère, bénéficient du hachage par le système FOIN pour les mêmes champs que ceux du SNIIR-AM, avec un système de clés conçu de telle sorte qu'il puisse y avoir rapprochement des données du PMSI avec celles du SNIIR-AM. Pour ce qui concerne les autres grandes bases de données nationales, comme par exemple les données de mortalité, la problématique est plus complexe car ces bases de données sont identifiées par le NIR du bénéficiaire et non pas celui de l'assuré. Mais on sait qu'avec la carte Vitale 2, le système d'identification du SNIIR-AM (et certainement du PMSI) doit très prochainement évoluer pour prendre en compte le NIR du bénéficiaire et non plus celui de l'assuré, ce qui permettrait de résoudre partiellement cette première difficulté. Il suffirait alors de procéder au hachage des trois traits d'identification utilisés dans la procédure d'anonymisation FOIN-SNIIR-AM (le NIR, la date de naissance et le sexe du patient), en utilisant le même système de clés, pour permettre l'appariement sécurisé (par exemple, par un tiers de confiance) des données du DMP avec les données de mortalité de l'Insee.

Toutefois, du fait que la procédure de hachage englobe les trois traits d'identification pour aboutir à un seul numéro, une simple erreur sur le sexe pourrait conduire à rejeter le chaînage des données d'un même patient. Il faut savoir également que le genre est une variable très peu fiable (de nombreuses erreurs sont observées sur le recueil de ce champ dans les établissements hospitaliers). Par ailleurs, sur le plan de la qualité du chaînage, cette variable est très peu discriminante dans la mesure où la probabilité que deux personnes aient le même sexe est de près de 50 % et ne permet, en particulier, pas de discriminer les jumeaux. Il serait donc préférable de remplacer cette variable par une autre variable beaucoup plus discriminante telle que le nom ou le prénom

1. Il faudra également prévoir l'anonymisation des professionnels de santé, de façon à pouvoir les rassurer sur les objectifs de l'utilisation épidémiologique des données : il ne s'agit en aucun cas de procéder à l'évaluation individuelle des pratiques professionnelles.




du bénéficiaire. Cela supposerait de modifier la procédure d'identification du SNIIR-AM, ce qui n'est pas forcément possible. Une autre solution serait de rajouter, à l'identifiant du SNIR-AM, les traits personnels du patient (pour être conforme aux recommandations de l'Association internationale d'informatique médicale, édictées lors de la dernière conférence — Dijon, avril 2006 — du Working Group 4 de l'IMIA tels que le nom, le prénom et la date de naissance du patient (hachés séparément). Cela permettrait aussi d'assurer l'interopérabilité de cet identifiant avec un identifiant européen, puisque le NIR ne peut servir de corrélateur entre un identifiant

français et un identifiant d'un autre pays d'Europe.

Conclusion

Afin d'assurer l'unicité de l'identifiant patient et sa compatibilité avec les principales bases de données de santé, l'utilisation du NIR semble souhaitable. L'anonymisation du NIR par une technique de hachage à sens unique devrait permettre de respecter les législations européenne et française sur le traitement des données à caractère personnel et de rassurer les associations de patients ou de défense des libertés individuelles.

Compte tenu de l'importance du choix et

de la nécessité de l'identifiant patient pour la réussite des enjeux du DMP, il nous paraît primordial que celui-ci préserve une interopérabilité future avec les dossiers patients des autres pays de l'Europe, et par voie de conséquence ne se limite pas à un dérivé anonyme du NIR mais englobe, après anonymisation, d'autres données d'identité. 

Références bibliographiques

1. Quantin C., Guinot C., Tursz A., Salomez J.-L., Rogier C., Salamon R. « Le traitement épidémiologique du Dossier médical personnel au service des malades ? » *Rev. épidémiol. santé publique*, 2006, 54 : 177-196.
2. Académie des sciences. *Rapport Science et Technologie*. Comité RST : « Épidémiologie : conditions de son développement, et rôle des mathématiques ». Animateur : Alain-Jacques Valleron. 11 janvier 2005.
3. Quantin C., Gouyon B., Allaert F.A., Dusserre L., Cohen O. « 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. [Consulté le 21/02/06]. URL : http://www.insee.fr/fr/ffc/docs_ffc/cs113e.pdf ; *Journal de la SFdS* 2005 ; 146 (3) : 15-26.
4. Roger-France F.H., De Clercq E., Bangels M. « Purposes of health identification cards in Belgium » In : Engelbrecht R., Geissbuhler A., Clovis C., Mihalas G., eds. *Connecting medical informatics and bio-informatics. Proceedings of MIE 2005*. Amsterdam : IOS Press, 2005.