Linkage of De-Identified Records in Accordance to the European Legislation

G

C Quantin
Service de Biostatistique et d'Informatique Médicale (DIM), France & INSERM U866 Université de Bourgogne, France

E Benzenine
Service de Biostatistique et d'Informatique Médicale (DIM), France

M Guesdon
Service de Biostatistique et d'Informatique Médicale (DIM), France

JB Gouyon
Centre d'Etudes Périnatales de l'Océan Indien, La Réunion, France

FA Allaert
The Evaluation of the Medical Claims of Health Foods, France

INTRODUCTION

Apart from uses required by health insurance or the authorities (electronic treatment forms, Programme de Médicalisation des Systèmes d'Information (Programme for the medicalisation of computer systems, PMSI) it is possible to consider using and sharing information gathered in medico administrative databases for statistical or epidemiological purposes. Nonetheless, gathering medical information relative to the same patient by linking various existing files must comply with European and French legislation relative to the protection of individual liberty. We will show that it is possible to gather together the different parts of a given patient's record without knowing the patient's identity, and thus comply with the above legislation by using cryptographic techniques, such as de-identification, and linkage procedures.

BACKGROUND

Legislation on the Security of Nominative Information in Europe

The European directive of 24 October 1995 (Directive 95/46/CE) relative to the protection of physical persons with regard to the processing of personal data and the free circulation of these data has defined "personal data," as "any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity." Given the above considerations, the notion of nominative or personal data concerns a vast amount of information even though the name no longer appears and there is no table to link the name of the person with the alphanumeric code that has replaced it. From a statistical point of view, the chances of identifying a person from apparently anonymous information is far from zero, given the possibility of linkage with the enormous diversity of existing or future files. Who in France would have believed, only a short time ago, that the French tax authorities would be authorized by law to computer process your National Health Insurance number?

Data processing was primarily regulated under the provisions for the Internal Market. Therefore, the general Data Protection Directive 95/46 combines two goals: protecting the fundamental right to data protection and ensuring the free flow of personal data within the internal market.

DOI: 10.4018/978-1-4666-5888-2.ch319

Most of the difficulties encountered in drafting the Directive, which aimed to harmonise protection provided by national legislation, came about because of the balance between these two principles.

Rather than opt for a maximalist solution that would have imposed as the standard the highest level of protection provided for in the different national legislations, the European legislator seems to have preferred a compromise, which was more difficult to establish, but complied more closely with the community view. It sought the highest level of protection that was acceptable without jeopardizing the protection of individual freedom and was yet compatible with the possible short-term evolution in the legislation in countries that provided the weakest protection to their residents. The resulting text, like all compromises, may have appeared to the most demanding countries as a step backwards with regard to the guarantees provided to citizens. However, it reflects the high priority the European Union gives to enabling less developed countries or countries with very different cultures to integrate the community process without too much antagonism.

Processing involving files that include personal data, and in particular health data, must be conducted in the context of a legal framework that guarantees the rights of patients (patients' information, data security and confidentiality).

Though the legal framework for processing health data in the context of healthcare or interventional research is clearly defined by current legislation, it is not the case when the final purpose of the processing is modified (reuse of health data for research purposes). Because of this legal void, the actors (healthcare professionals, researchers, managers of healthcare information systems) have to identify the ethical and legal aspects to respect in this new context.

In France, the French Commission for Data Protection (CNIL) and various Committees for the Protection of Persons (CPP) have put forward several proposals to establish an agreement protocol concerning the sharing of data within hospitals. Certain hospitals implemented, following approval from the CNIL and the CPP, a health-data sharing surveillance committee within the establishment. The role of this committee is to provide information on and apply the current legislation by approving and authorizing such uses.

METHODS

Anonymity: The Application of Cryptographic Methods

Statistical methods that depend on the disruption of data to render them anonymous lead to reduced information quality. It therefore seems preferable to use cryptographic techniques to ensure information security. These techniques, which make it possible to protect information thanks to a secret key, generally stem from mathematical problems which are extremely difficult to solve without the key. They are not particularly recent, but their use was restricted by French law for reasons of national defence. The field of cryptography benefited from a certain degree of liberalisation, first of all in 1998 when the procedure for declaration to the French National Agency for the Security of Information Systems (Agence Nationale de la Sécurité des Systèmes d'Information, ANSSI) was simplified, then in 1999, placing the burden of the legislation on cryptology professionals. This evolution was made possible in particular by the use of high-security keys 128 bits long (previously limited to 40 bits). Such keys became compulsory to meet the requirements for electronic signature recognition and to facilitate commercial transactions on the Internet. This liberalisation made it possible to overcome obstacles that prevented the use of cryptographic techniques to ensure the confidentiality of directly or indirectly nominative medical information likely to circulate on computer networks. With regard to the security of medical information circulating on a network, the encryption methods can be used at three levels.

The first level consists in ensuring the confidentiality of a communication between two persons. The second level is to verify that a person is the author of a message by using an electronic signature. The third level of cryptographic techniques concerns the gathering of medical information within a structure outside the establishment where the patient was treated. This is the level we are interested in.

Hashing Techniques

Hashing consists in applying a hash function on data. Such a function takes any input data and maps it to a fixed length bit vector, sometimes called a signature.