Chapter 11
Proposal for Interactive Anonymization of Electronic Medical Records

Carlos Andrés Moque Millán  
Pontificia Universidad Javeriana, Colombia

Alexandra Pomares Quimbaya  
Pontificia Universidad Javeriana, Colombia

Rafael A. González  
Pontificia Universidad Javeriana, Colombia

ABSTRACT
One of the most important inputs for medical research is the information registered in electronic medical records. This information typically contains sensitive data that must be preserved in order to be used for research or educational purposes, and protected depending on the regulations of each country and institution. In order to assure confidentiality of data, different techniques can be used to remove basic identifiers (e.g. names, IDs); however, these techniques can be easily bypassed by attackers who know the information that can act as pseudo-identifiers of patients (e.g. birthdates, gender). Although these pseudo-identifiers can also be removed, the information they contain is valuable for medical research. To face this problem, different methods that allow minimizing the risk of sharing confidential information have been proposed. The interactive use of anonymization algorithms for electronic medical records is the main contribution of this chapter, dubbed AnonymousData.co: a proposal for anonymization of electronic health records.

DOI: 10.4018/978-1-4666-3667-5.ch011
INTRODUCTION

Manipulating sensitive data from patients for medical research generates a clear threat to privacy and data protection. Anonymization of such data is thus a known strategy for minimizing this risk as well as preventing misuse by malignant attackers, while at the same time preserving the consistency and value in the data. In this chapter, the data in question is composed of electronic medical records stored in the information system of the San Ignacio University Hospital in Bogotá, Colombia. This data is being treated with data mining techniques in order to identify patients that have been diagnosed with a specific chronic disease, in order to aid with clinical research as well as identifying patterns for analyzing adherence to diagnosis and treatment standards.

According to Article 15 of the Colombian Constitution, all citizens have the right to preserve their personal and family intimacy, as well as their good name. With the growth of the “information society” it has become increasingly difficult to guarantee this right, as personal data is being stored in an ever increasing number of databases owned by both private and public institutions. Accordingly, it has become more critical to adopt strategies for encoding or removing sensitive data if and when personal records are going to be used for analysis, achieving a sufficient balance between anonymization and usefulness of the data in order to generate meaningful knowledge (Aggarwal & Yu, 2008). Striking this balance is a political, social and technological problem which should always aim at generating value for the citizens themselves without trampling their right to privacy.

From a technological perspective, there is a number of available anonymization strategies focused on microdata. However, and especially after the advent of the Internet, it has been possible to re-identify private data by combining public data and microdata through attributes that serve as quasi-identifiers (QI), as defined in (Samarati, 1998). QI represent those attributes that do not directly identify an individual but which may be grouped to decode sensitive or confidential information.

This work will proceed by exploring some of the most used anonymization methods applicable to electronic medical records. It will present the results of applying those methods to a subset of records from the aforementioned hospital information system. The main goal is to guarantee, following (Friedman, 2008), that privacy is protected, but that the data remain rich enough to carry out a data mining exercise.

The rest of this chapter is structured as follows. Section 2 summarizes the conceptual foundations of this work in terms of anonymization methods and models. Section 3 presents the anonymization proposal, centered on microdata and emphasizing the role of the interactive anonymization process for the hospital case. Finally, Section 4 concludes this paper and links it to the next steps in the hospital project.

CONCEPTUAL BACKGROUND OF ANONYMIZATION

Before presenting some of the available anonymization alternatives, it is useful to define some relevant concepts used throughout the chapter. Sensitive data refers to private information which must be protected in order to guarantee confidentiality and which may be misused with harmful purposes – for example, a chronic disease diagnosis for a given person. Quasi-Identifiers (QI) are attributes (from a database table) which, when grouped together, may become identifiers that reveal sensitive data – for example, while a national ID number unequivocally identifies an individual, it may also be possible to identify him or her if you know the age, gender and approximate address. An attacker is an individual which misuses QIs in order to extract sensitive information about a person or group with the intent to harm or benefit from this data – for example, a
Related Content

A Comprehensive Overview of Wireless Body Area Networks (WBAN)
[www.igi-global.com/article/comprehensive-overview-wireless-body-area/55998?camid=4v1a](www.igi-global.com/article/comprehensive-overview-wireless-body-area/55998?camid=4v1a)

Relationship Between Performance Error and Human Information Processing
[www.igi-global.com/chapter/relationship-between-performance-error-human/13059?camid=4v1a](www.igi-global.com/chapter/relationship-between-performance-error-human/13059?camid=4v1a)

Rescheduling Dental Care with GSM-Based Text Messages
[www.igi-global.com/chapter/rescheduling-dental-care-gsm-based/31164?camid=4v1a](www.igi-global.com/chapter/rescheduling-dental-care-gsm-based/31164?camid=4v1a)

The Vascular System: Unblocking Conduits to Our Rivers Within
[www.igi-global.com/article/the-vascular-system/115532?camid=4v1a](www.igi-global.com/article/the-vascular-system/115532?camid=4v1a)