A Web Architecture Based on Physical Data Separation Supporting Privacy Protection in Medical Research

Jan Sliwa, Division of Computer Science, Bern University of Applied Sciences, Bern, Switzerland

Emmanuel Benoist, Division of Computer Science, Bern University of Applied Sciences, Bern, Switzerland

ABSTRACT

In this paper, the authors present a novel Web-based architecture of a medical registry with enhanced protection of personal data of the patients. The goal of a medical registry is to gather experience from clinical practice concerning a disease or treatment (e.g., hip replacement) and to improve the future treatments by applying adequate methods and selecting optimal products. The processing of health data is strictly regulated by laws protecting patients’ privacy. The presented solution is based on the physical separation of identity related data and clinical data and combining the information from both sources on the screen of the user, according to their permissions. The anonymized clinical data can be used for research whereas the risk of de-anonymizing the patient is significantly reduced. This solution has been verified by an operation of a deployed real-life application.

Keywords: Anonymization, Medical Documentation, Privacy Enhancing Technology (PET), Privacy Protection, Quality Assessment, Web-Based Architecture

INTRODUCTION

Medical registries provide a platform for collecting the outcomes of medical treatments in a systematic way. Recording the initial conditions of the patients, the applied treatments and their results, short- and long-term, permits assessment of the relative quality of various methods and devices and to determine their optimal scope of application. This should lead to better decisions in the future and in effect to an improvement of patient care and to a cost reduction. The value of a registry increases with the quantity and quality of collected data. On the other hand, these data are personally highly sensitive, possibly stigmatizing. When disclosed, they may be used as a basis for discrimination and profiling. Their privacy is strictly regulated. However, when subsequent
data records related to a patient are entered, e.g., for a follow-up examination, he/she has to be retrieved, therefore his/her real identity must be stored in the system. For the purpose of medical research these identities are irrelevant but the data set of a patient must still be connected. These requirements are partly contradictory but we have to satisfy them all. In this paper, we present a novel architecture based on the physical separation of the identity related data and the anonymized clinical data that greatly limits the risk of a privacy breach while still allowing for valuable medical research. We have discussed these trade-offs with a special stress on the ethical aspects in Sliwa and Benoist (2011).

We work in a multifaceted area concerning the following aspects:

- Privacy / security technology,
- Utility requirements of medical registries,
- Compliance to privacy legislation.

First, we present the rationale for medical registries and their organization as statistical databases. We also discuss the legal environment in which these registries operate. Then we present the architecture of our solution. We give examples of the problems related to the separation of the data categories and show their solutions. Finally, we summarize the experience gained from our project and outline the possible future work in this area.

BACKGROUND

Medical Registries

Novel medical treatments and new medical devices need exhaustive quality control. In order to draw meaningful conclusions and to take reasonable decisions, a comprehensive data set about the preliminary health conditions, applied treatments and their outcomes is necessary. We may divide the collections of medical outcome data in two basic groups. The first group regards the clinical trials. Their goal is to verify a clinical question, clearly stated beforehand. The patient group and the time period of the study are also well defined. The second group comprises the registries where we gather data concerning a certain disease and/or treatment and we do not have a specified hypothesis to verify. Therefore it is difficult to apply the rule of gathering only the data necessary for a defined purpose because a new interesting hypothesis about dependencies and correlations might be formulated also later on. We have to remember that obtaining every data set has a cost; it corresponds to a real operation of a patient and missing or forgotten data cannot be completed retroactively.

In the following text, we will concentrate on registries. The complex subject of defining and running registries is presented extensively in Gliklich and Dreyer (2010). A classical solution is the Swedish Total Hip Arthroplasty (THA) Register (http://www.jru.orthop.gu.se/), established in 1979, containing about 315,000 primary operations (2009), with some cases followed for decades. Another important project is the International Spine Registry (SpineTango, http://www.eurospine.org/p31000381.html) implemented under the auspices of the Spine Society of Europe (EuroSpine) and collecting data from the clinics all over the world.

The goal of a registry is of course not just collecting data. The rationale for a spine registry is well presented in Röder, Müller, and Aebi (2006). A team of qualified specialists in medicine and statistics can draw conclusions about the quality of used methods and devices and formulate recommendations for future treatments, solidly based on past experience and supported by verifiable statistics. Examples of such analyses based on the hip and spine registry can be found in Röder, Staub, Eggli, Dietrich, Busato, and Müller (2007); Röder, Bach, Berry, Eggli, and Ronny Langenhahn (2010); Diel, Reuss, Aghayev, Moulin, and Röder (2010). The knowledge gained should lead to applying a treatment optimally matching the patient’s case, lowering the risk of a re-operation and reducing the health care costs.