Collating of a Distributed XML Based Medical Records into a Relational Database

Do Hoon Oh MD, Alberto Riva PhD, Kenneth D. Mandl MD, MPH, Isaac S. Kohane MD, PhD

Children's Hospital Informatics Program and Division of Endocrinology
Children's Hospital, Boston, MA

Introduction

We are currently developing the Personal Inter-networked Notary and Guardian (PING)\(^1\) whose main goal is to implement a secure, distributed and patient-controlled repository of life-long medical information on the Web. The PING record differs substantially from institutionally-oriented medical information systems. First, it uses XML as a basic data format and virtual directory structure as data storage and does not employ a relational database management system. Second, it is a highly secure system under a patient’s own control on the bases of authentication, authorization, and encryption technology. Third, PING records are distributed over the Internet and not confined to a single institution’s database. Fourth, PING records may have multiple data sources including patients as well as several health care providers. PING records could therefore provide very extensive, comprehensive and valuable information that other medical information systems may not encompass. However the essential distributed nature of PING records imposes several functional limitations in their use as a data warehouse for public health research. Consequently, we have addressed the construction of an efficient data management system required to generate useful public health information from PING records. We have developed a PING Poller, which retrieves PING records via a PING server and parses, reorganizes these data and inserts them into a relational database management system. This collation system, across widely distributed health records, will allow health care professionals and public health researchers to access PING records for clinical, epidemiological and public health researches, if given appropriate authorization by patients.

Implementation

The PING Poller is currently implemented as a Java application (using the Java 2 virtual machine) that broadcasts queries to each of a list of PING records (each representing a patient) for a particular data subset (e.g. immunization history). The PING Poller’s target is the data residing on encrypted, XML-formatted records stored on generic web server. As these PING records are encrypted, all access occurs via a PING server. Therefore, the PING Poller has to pass a cryptographic authentication test of the PING server and furthermore, the patient must have authorized release of the information. Patient-initiated release/authorization is effected by the patient setting the appropriate permissions on the hierarchical file structure of their PING record (via a user interface with the PING server). Authentication of the PING Poller is effected by using a cryptographic certificate recognized/registered by the PING server.

Upon receiving a requests PING server sends PING data as XML documents. The PING Poller receives and parses these documents and reorganizes data into a format suitable for insertion into a relational database. The PING Poller makes its connection to the database system via the JDBC protocol and inserts each patient record into the corresponding table. This table can then be queried by the investigator/public health authority with the SQL data manipulation language.

At present PING Poller has been only implemented for infants immunization history and allergies. These data types have been chosen as the initial test set as they carry relatively low-sensitivity information which is nonetheless of significant clinical import.

References