Data Repository for Clinical Research

1 Peisan LEE, 2 Yu-Chuan LI & 2 Yung-Tai YEN

1 National Yang Ming University, TW
2 Taipei Medical University, TW

Researchers reveal that an integrated, multi-central, longitudinal, and research-oriented clinical data repository (ROCDR) transformed from real patients' Electronic Health Records (EHRs) can be of immense value for clinical research on etiology, diagnosis, treatment, prognosis, and management of diseases as well as for the research on the quality and safety of care. A pool of the patient data matching a specific profile can be valuable for the clinical research. Unfortunately, they are often very difficult to obtain from even a single healthcare organization. The complicated processes of gathering electronically available patient data residing in multiple centres have prevented most researchers from seemingly significant and large-scale discoveries in biomedical informatics. The obvious benefit of Electronic Health Record (EHR) in clinical research seldom materializes in a multi-centre setting. To address this issue, we started a pilot project with four medical centres in Taiwan. All the participants agreed to contribute for 3 years of de-identified patient data into a distributed repository with a web-based user portal for data retrieval.

This research-oriented clinical data repository features a set of standard data formats not only for diagnoses, patient profiles, medications, examinations, laboratory data but also procedures obtained from different departments like outpatient, inpatient, and emergency. The architecture of ROCDR was designed as a naturally scalable grid-computational environment for end users (researchers). There is one database server hosted in each hospital to store patient data, therefore data access can be fully controlled by hospital administrators. A web-based user portal was developed to generate queries and retrieve data from each database server concurrently and it is easy to add new hospitals in this architecture. The web service is deployed in the service centre to accept information requirement; the following services will be deployed in the ROCDR: 1) data retrieval and calculation service: 2) full query resulted data will compress for downloading as txt file after IRB approval.
Patient IDs were encrypted using a commutative hash function to mitigate the threat of invasion of individual privacy and ROCDR was approved by the Institutional Review Board (IRB) of each participating hospital. In this article, we will describe how to make clinical data available and show the naturally scalable architecture of ROCDR.