Proceedings

SY03.03 | Data Integration and Modelling

Data Architecture For A Clinical Data Repository - Evaluation And Design At Charité Universitätsmedizin Berlin

M. Mallach*, M. Peuker
Charité Universitätsmedizin Berlin, IT Department, Berlin, Germany

Introduction/ Background

The translation of scientific results into new and more effective diagnostic and therapeutic procedures is a milestone in the advancement of medicine. German clinics collect large amounts of data on every patient, which is used to: evaluate treatment; justify expense reports; operate quality assurance and to keep aftercare Physicians/caregiver informed. However, so far there are only fragmentary approaches to using this data resource. The combination of phenotype information from clinical routines with information about samples held in the biobank systems and linked with genetic information must be achievable. Therefore a data architecture must be designed which allows an access of all relevant patient data stored in different source systems under the aspects of data security, data protection, data integrity and semantic interoperability. The following paper focused on the integration of patient clinical data and patient sample data stored in the biobank system. The central role plays the design and implementation of a Clinical Data Repository (CDR) at the Charité.

Aims

Definition of the essential requirements of a Clinical Data Repository regarding data security, data protection, data integrity and semantic interoperability. Definition of standardized data flow from clinical patient system into the clinical data repository. Definition of the data model of the clinical data repository.

Methods

In order to design the central Clinical Data repository an investigation of existing clinical and research system landscape at the Charité took place. The existing solutions were evaluated regarding usage, level of penetration, standards in interoperability and supported interfaces. The analysis of Clinical Data Repository requirements was based on the methodology of Requirements-Engineering. This methodology is very often used for development of complex IT systems in order to gain a common understanding on user side and developer side too. In 2013 and 2014 Charité supported a project to identify the main system demands on a clinical scientist workplace. The essential demands incorporated the Clinical Data Repository requirement analysis. From the technically point of view the Clinical Data Repository has to deal with a large amount of data in terms of storage, scalability, stability, fast accessibility and search and the support of data analytics. The introduction of In-Memory technology has enabled a paradigm shift in analytic applications, with many new possibilities. It is now possible to have all working data in the main memory, which means that internal database programming can be implemented to execute computer and data intensive algorithms without having to access data over slow interfaces. The Clinical Data Repository will be based on latest In-Memory technologies to allow researchers and physicians real time access to huge amounts of data.

Results

In the first phase the design of data architecture and a core set of clinical date is defined. A pilot implementation of a Clinical Data Repository will provide a research space where clinical data and research data is accessible for researchers and physicians. To correspond with data protection laws this data will be
anonymized, pseudonymized and stored securely. The central biobank system is connected via an identify management system with the Clinical Data Repository. The CDR allows requests to identify groups of patient with include or exclude criteria from clinical workplace as well from biobank system.