Article published online: 2022-12-04

Appendix: Summary of Best Papers Selected for the 2021 Edition of the IMIA Yearbook, Clinical Research Informatics Section

Bahmani A, Alavi A, Buergel T, Upadhyayula S, Wang Q, Ananthakrishnan SK, Alavi A, Celis D, Gillespie D, Young G, Xing Z, Nguyen MHH, Haque A, Mathur A, Payne J, Mazaheri G, Li JK, Kotipalli P, Liao L, Bhasin R, Cha K, Rolnik B, Celli A, Dagan-Rosenfeld O, Higgs E, Zhou W, Berry CL, Van Winkle KG, Contrepois K, Ray U, Bettinger K, Datta S, Li X, Snyder MP

A scalable, secure, and interoperable platform for deep data-driven health management

Nat Commun 2021 Oct 1;12(1):5757

This paper presents a major effort in building a secured and scalable platform to gather big biomedical data from different sources (including genomics, EHRs, wearable sensors). The authors focus on technical aspects of building such a platform: security (local data storage, defense against reverse engineering, mobile app security, anonymization), scalability (authentication, messaging, machine learning, infrastructure based on the open-source tool Terraform) and analysis (data preprocessing, feature extraction). Although interoperability issues are not really mentioned, several APIs for data collection are described. The main features of the platform are data visualization, monitoring and alerts, as well as feature prediction with logistic regression (84 features). The platform can be used at patient-level or at cohort-level and has been used for the detection of pre-symptomatic COVID-19 cases, and for biological characterization of insulin-resistance heterogeneity. In conclusion, this very large-scale platform for biomedical data offers guarantees in terms of security, scalability, data preprocessing and provides features for visualization, monitoring and analysis.

Cheng AC, Duda SN, Taylor R, Delacqua F, Lewis AA, Bosler T, Johnson KB, Harris PA

REDCap on FHIR: Clinical Data Interoperability Services

J Biomed Inform 2021 Sep;121:103871

This paper describes the development and evaluation of the REDCap Clinical Data Interoperability Services (CDIS) module that provides seamless data exchange between the REDCap research Electronic Data Capture (EDC) system and any EHR system with a FHIR API without project-by-project involvement from Health Information Technology staff. An iterative process has been used to design all aspects of the CDIS module (access control, authentication, variable selection, and mapping) in such a way that end users could easily set up and use the module in 2 use cases. In the "Clinical Data Pull" (CDP) mode the CDIS automatically pulls EHR data into user-defined REDCap fields. In the "Clinical Data Mart (CDM)" mode, the CDIS collects all specified data for a patient over a given time. Beyond the stakeholders group initially involved including Vanderbilt University Medical Center (VUMC) health IT and EPIC EHR teams, other healthcare organizations and EHR vendors have been associated through the REDCap consortium. As of Nov 2020, since its release (1st CDP project live@ VUMC launched Q3 2018, REDCap released on Epic App Orchard Q1 2019), 82 projects are running at VUMC (55 CDP, 27 CDM) with 19.5 M data points transferred. With a large scale adoption in REDCap consortium sites (26 implementations in other institutions with EPIC EHRs / 47 ongoing implementations in institutions with EPIC (n=26) or Cerner (n=9) EHRs), the REDCap Clinical Data and Interoperability Services (CDIS) are key contributions to the integration of care and research activities. Thanks to the CDIS modules, leveraging the FHIR standard to use of EHR as electronic source for clinical research, the researchers can self-service the setup of real time and direct data extraction from the EHR reducing the need for manual transcription and flat file uploads and improving the accuracy and efficiency of EHR data collection.

Pedrera-Jiménez M, García-Barrio N, Cruz-Rojo J, Terriza-Torres AI, López-Jiménez EA, Calvo-Boyero F, Jiménez-Cerezo MJ, Blanco-Martínez AJ, Roig-Domínguez G, Cruz-Bermúdez JL, Bernal-Sobrino JL, Serrano-Balazote P, Muñoz-Carrero A

Obtaining EHR derived datasets for COVID-19 research within a short time: a flexible methodology based on Detailed Clinical Models

J Biomed Inform 2021 Mar;115:103697

Responding to the urgent need for health data insights during the COVID-19 pandemic, and utilizing this as a use case for a generalizable methodology, Pedrera-Jiménez et al report on the use of Detailed Clinical Models (DCMs) as a formalized representation for the structure and semantics of research data sets. They propose this as a data transformation pathway to generate datasets rapidly and accurately from EHRs for secondary use, without loss of meaning or error, allowing for frequent changes in specification, and being easy to validate.

The authors took as their use case the need to rapidly generate a research data set conforming to the International Severe Acute Respiratory and emerging Infection Consortium (ISARIC-WHO) COVID-19 specification. Instead of the classical approach of authoring this data set in an electronic case report form (eCRF), the authors modelled this research data set as a portfolio of Detailed Clinical Models: EHR archetypes conforming to the ISO 13606 standard. These archetypes each expressed a specific data structure pattern which was a profiled subset of the generic 13,606 EHR interoperability reference model, and incorporated semantic constraints (e.g,. value sets) drawn from SNOMED-CT or LOINC, as appropriate for clinical or laboratory concepts. These DCMs were used as the data extraction mapping target from the EHR system at the Hospital Universitario 12 de Octubre in Madrid. The extraction included data on 4,489 patients hospitalised with COVID-19 over a six-month period during 2020. The flexibility and agility of this method was demonstrated

through the ability to revise the data set specification easily by modifying or adding further archetypes. The authors discuss the future potential of this method to also utilise HL7 FHIR resources as alternative DCMs, through a forthcoming ISO Technical Specification on "Guidelines for implementation of HL7/FHIR based on ISO 13940 and ISO 13606". There is a growing need and opportunity for the systematic and interoperable reuse of routinely collected (real-world) EHR data for research, through the centralised or federated querying of standardised data sets. This research, although mono-centric and exemplified through COVID-19, is included as a best paper because the methodology is generalisable to any area of clinical research for which there is relevant real-world data.