Cross4all Project Model of Integration of Healthcare Data Using the Concepts of EHR and PHR in the Era of IoT

Andrijana Bocevska¹, Snezana Savoska¹, Blagoj Ristevski¹,
Natasha Blazheska-Tabakovska¹, Ilija Jolevski¹, Vladimir Trajkovik²

¹Faculty of Information and Communication Technologies – Bitola,
University “St. Kliment Ohridski” University – Bitola, ul. Studentska bb 7000 Bitola
Republic of North Macedonia
²Faculty of Computer Science and Engineering, “Ss. Cyril and Methodius” University Skopje
Republic of North Macedonia
¹{andrijana.bocevska, snezana.savoska, blagoj.ristevski, natasa.
tabakovska, ilija.jolevski}@uklo.edu.mk
²vladimir.trajkovik@finki.ukim.mk

Abstract. Many efforts are made to integrate healthcare data, hospital information systems data, clinical and medical data to provide healthcare data analysis to be suitable for healthcare decision-makers. All these heterogeneous data are stored in many different places, formats and heterogeneous platforms and their integration is a very challenging and demanding task and sometimes even impossible. The existence of patients’ related healthcare data issues is evident, although they are stored in various hospital and public health systems such as Electronic Health Records, healthcare institutions and laboratories, patient’s health records, medical records. In this paper, we describe the Cross4all project model of integration of healthcare data into Personal Health Records with a focus on the patient, into the cloud environment with required data security and privacy standards.

Keywords: Health Data Integration, Electronic Health Records, Personal Health Records, Wearables.

1 Introduction

Today’s trends of collecting data from healthcare and medical issues, from different sources as hospital information systems (HISs), Electronic Health Records (EHRs), medical prescription, diseases diagnoses and treatments demand a serious approach to an ontology-based collection of data according to HL7 standard and healthcare data security and privacy demands. Healthcare data are owned by many healthcare providers and are not accessible to patients. Data can be depersonalized and only used by decision and policymakers if they are

Copyright © 2021 for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).
integrated. A lot of research is dedicated to depersonalized medical and health data integration intended for the municipality, state and hospital management structures. If we take into consideration data collected from many wearables that support the ambient assisted living (AAL) concepts and help to improve medical and healthcare, the integration of patient’s healthcare data is very important. Many heterogeneous data in the personal health record are added from the sensors as part of the Internet of Medical Things (IoMT) concept, trackers for human behavior and vital signs of life, as well as exposome data.

Strong security standards for healthcare and medical data for patient’s centric systems with personal health record (PHR) are implemented regarding the secured share of healthcare data with a temporarily selected doctor. This concept demands a complex cloud infrastructure, with security and data protection procedures, made according to the national protection law. The standards as HL7, FHIR, open EHR and coding systems as ICD10 are also necessary to be used.

The paper is structured as follows. The second section provides related works, whereas the third section highlights the patients’ data privacy toward securing their electronic PHR (EPHR) data, as well as the reusability of EPHR data, usage possibility and usage disadvantages. In the subsequent section, we describe the Cross4all project model of integration of healthcare data into PHR, providing practical examples of model implementation in the project activities. The final section gives concluding remarks and points out some directions for future works.

2 Related works

Many efforts were made by researchers in the health and medical domain to provide reusing of healthcare data. A model for decision makers’ support according to the national law, using Spark, Mongo DB and DL-bases AI module for NLP is proposed in [1]. The interoperability among EHR systems, the full integration of clinical data within PHRs and the exploitation of the contained information is a widespread target internationally.

An open data integration platform for patient, clinical, medical and historical data, siloed across multiple HISs is proposed in [2]. The platform was adopted and implemented to address patient-centered healthcare and clinical decision support requirements in a sports injury clinic at a not-for-profit private hospital in Australia. It can accommodate and integrate further heterogeneous data sources such as data streams generated by wearable IoT devices. The distribution of scanned documents at one health institution and the design and evaluation of a system to categorize documents into clinically relevant and non-clinically relevant categories as well as further sub-classifications were described in [3].
A method for digitizing the concept of health by processing the existing information in EHRs with the help of several dedicated services was presented in [4]. It introduces the “health digital state” (HDS) as a digital equivalent to the “health” concept. A business use case that is extremely common in current medicine: the encounter between a patient and a healthcare professional caused by the worsening of the patient’s health is implemented to explain the concept of HDS and its use in an advanced EHR system. Precision Medicine includes the discovery of a patient-specific pattern of disease progression, as well as a determination of the precise therapy for that pattern, and the corresponding personalized delivery of care [5]. Although EHRs are instrumental across this spectrum, they focus on personalized healthcare delivery based on the rapidly evolving knowledge base brought about by advances in genomic medicine.

An Internet of Medical Things (IoMT) platform for pervasive healthcare that ensures interoperability, quality of the detection process, and scalability in a machine-to-machine-based architecture and provides functionalities for the processing of high volumes of data, knowledge extraction, and common healthcare services, was proposed in [6]. The platform uses the semantics described in OpenEHR for both data quality evaluation and standardization of healthcare data stored by the association of IoMT devices and observations defined in OpenEHR. In [7], the authors had demonstrated the feasibility of a scalable, accurate, and efficient approach for medical device surveillance using EHRs. They presented that implant manufacturer and model, implant-related complications, as well as mentions of post-implant pain can be reliably identified from clinical notes in the EHR.

Three threats from real cloud-based electronic healthcare (eHealth) systems, i.e., privacy leakage, frequency analysis, and identical data inference had been identified in [8]. They propose a multi-source order-preserving encryption (MSOPE) scheme for cloud-based eHealth systems, which enables doctors to perform privacy-preserving range queries over encrypted EHRs from multiple patients.

The authors in [9] built an EHR Aggregator (EHRagg) that integrates the developments made so far to learn automatically how to convert current information systems into standard systems. With the EHRagg they address the interoperability and accessibility problem using the same pragmatic approach: instead of trying to have all the systems agree with the same standard, they propose a translation between standards, and of systems to any standard, reducing effort and time. Wireless sensors in the IoT context in contemplation of model solutions in the field of eHealth were investigated in [10]. The focus of their work is on merging the person’s health data collected through wearable and non-wearable sensors into the formal infrastructure and services within Croatia’s central health information system. The process encompasses a collection of data and transform-
ing the data collected into a proper medical format (HL7 or FHIR) to ensure the data is structured and easy to understand.

The authors in [11] outlined various secondary uses of EHR to give an idea of how effectively EHR data can be used in different domains such as clinical research, public health surveillance and clinical audits to provide effective, timely and quality healthcare facilities to the patients. Data security and patients’ privacy risks related to the secondary uses of EHR especially when EHR data are transmitted through a network and shared with multiple stakeholders are also critically studied. Different database models’ appropriateness for integrating different EHRs functions with different database specifications and workload scenarios were discussed in [12]. According to their related works’ analysis, every database technology offers diverse health care task performance according to this database’s specification and related workload types.

The authors in [13] describe steps necessary to use the EHR as a tool for conducting high-quality clinical research. They mention the inadequate or complete lack of standard data structures in current EHR as a problem in using point-of-care data for research and examine the changes necessary for reconfiguring current electronic health records to collect data of sufficient quality to support the most stringent research methods, namely randomized clinical trials (RCTs).

Applications of unsupervised machine learning approach discovering latent disease clusters and patient subgroups using EHR data were described in [14]. They utilized Latent Dirichlet Allocation (LDA), an unsupervised probabilistic generative model in the rubric of topic models, and proposed a novel unsupervised machine learning approach Poisson Dirichlet Model (PDM).

A knowledge-driven framework able to transform disparate data into knowledge from which actions can be taken to help clinicians and data practitioners in the complex tasks of extracting valuable knowledge from heterogeneous datasets is described in [15]. They describe the application of the framework in the biomedical domain and show the potential for uncovering patterns that can enable the explanation of treatment interactions and patient characterization.

A tethered PHR that seeks to achieve interoperability by using open-source standards and their implementation is presented in [16]. A tethered PHR application achieves both structural and semantic interoperability to allow data exchange with external systems such as an EHR easing data integration issues and improving data quality. The prototyped mobile PHR uses the guidelines narrated in the HL7 PHR-S FM for its functional requirements, the new HL7 FHIR for capturing and sharing data and SNOMED for attaching semantics to the captured data. The primary goal of the prototype is to demonstrate the capability of HL7 FHIR and its features (profile, extensions, and capability standard) to design and implement an interoperable PHR that aligns with HL7 PHR-S FM. As HL7 FHIR is a specification, the EHR and mobile PHR leverage the HAPI FHIR, a Java
implementation of the HL7 FHIR. The data captured in the PHR is structured as FHIR resources and shared in JSON format with the EHR using web services. According to the conclusion in [17], cancer genomic information integration into EHRs could help promote the benefits of patient-centered care. Machine learning algorithms and CDS software will harness cancer genomic-EHR integration. They suggest clinicians to be more inclined to let the genomic information in their patients’ EHRs better guide the decisions they make if it is well integrated.

Several EU projects intend to focus healthcare data integration on the patient, providing patient-centric healthcare data integration cross border through PHR where the patient is the data owner [18]. The security issues are considered from the aspect of the patient and living country of the patient [19]. The proposed model is cloud-based cross-border healthcare system based on the PHR concept with an e-health strategy. The key point is that data collection can be made sometimes out of hospitals and HIS and perhaps it cannot be connected with EHR and country of living. This concept demands increasing the e-health and health digital literacy to be implemented as well as the support of the national and local medical and healthcare authorities [18] [20].

Some authors think that healthcare data integration has to be wider and has to provide wider data integration, for not only data analysis and healthcare decision-making. An integration of healthcare data, medical, omics, sensors data as well as exposome data to provide data for prediction of the influence of health of environmental, social, stress factors to risk to health assessment was proposed by [21] and [22].

The Cross4all project model integrates healthcare and medical data from different sources as EHR, HISs and measurement sensors into PHR as the first stage towards the integration of patient health data. These data, as well as numerous biological omics and exposome data and data obtained from wearables, are considered and stored on the cloud following the required data security and privacy standards.

3 Privacy and personal health data of patients

Providing patients with proper healthcare information and health facilities at a low cost has always been a great challenge for health service providers. It includes health monitoring in- and out-of-hospital conditions for older people and patients who need supervision. Recent advances in wireless sensor technology envisage new types of ubiquitous healthcare systems [23, 24]. These systems provide permanent monitoring of patients, even during their normal daily activities, without compromising their quality of life, enabling the development of patient-centric pervasive environments in addition to the hospital-centric ones [19]. Such systems will enable healthcare personnel to timely access, review, and update
and send patient EHR from wherever they are, whenever they want [21]. Some of those systems are based on open platform interfacing to a wide variety of sensors, collecting and storing the data in a server repository, and making the available EHR data applications through a documented API [25].

There are several architectural models for building this kind of personal healthcare system [26]. Pervasive health provides technologies that help citizens participate more closely in their healthcare [24]. They should provide flexibility in patients who lead an active everyday life with work, family, and friends [27]. However, these systems do not consider the collaborative value that can be provided with matching gathered data. Dataflow of patients with the same diagnoses can be provided directly or as grouped statistical summaries. It enables the exchange of patients’ experiences by using the activities that other patients have taken.

In the process of designing a healthcare system, the following points about Electronic PHR data should be considered:

1. Data needs to be actionable to patients, policymakers, and health care providers in order to help them make better decisions.
2. “Everyday” data needs to be considered and can be as valuable as lab tests in its impact on their health outcomes.
3. Patients and healthcare providers need to look at their relationship as a collaboration, which requires a new definition of the doctor-patient relationship.
4. Technology offers opportunities, but it is not the silver bullet. It cannot be intrusive; it needs to be a part of an individual’s life.

A simple overview of the typical distributed healthcare system model is presented in Fig. 1. The system is deployed over three primary pillars:

1. The first pillar consists of the bio network (implemented from various body sensors) and mobile application that collects users’ biodata during various physical activities (e.g., walking, running, and cycling).
2. The second pillar is presented by the social network implemented as a web portal, enabling different collaboration within the end-user community.
3. The third pillar enables interoperability with the primary/secondary health care information systems, which can be implemented in the clinical centers and different policymaker institutions.

Communication between the first and third pillars of the model is determined by communication between patients and healthcare centers. The patient has 24-hour access to medical personnel with the possibility to make an emergency call. The medical staff monitors the patient’s medical condition remotely, reviewing the medical data and responding to the patient by suggesting the most suitable therapy. The medical personnel can also send patients various notifications (e.g., tips and suggestions) regarding his/her health condition.
The second and third pillars can exchange data and information regarding the larger group of patients by any significant indicator (region, time, sex, type of activities), which can be later used for research, policy recommendations, and medical campaign suggestions.

One of the most critical issues in the system is information validity and confirmation. We can divide system information validity into three categories [28]. Most reliable information is information that originates from the medical databases, clinical centers, and sometimes biosensors depending on their usage factors. Less reliable information is information generated from Social networks. This information can increase its reliability if confirmed by clinical centers’ medical records. Information from personal profiles (age, weight, height, diagnosis entered by end-user) is the third category of information (unreliable information). An increase of validation of this information can be done by comparing them with average results using a social network or by confirming them with the medical records coming from healthcare institutions.

![Fig. 1. Distributed healthcare system pillars.](image)

The categorization of the validity of the information can be used to determine the validity of notifications created within the system. It is essential because it affects the users’ decision whether to respect the notification or not. Every user can determine what information can be private or public. To obtain medical support, the user has to agree to share personal information with clinical centers and medical databases, whose data are protected. According to the user agreement policy, data information would be exchanged through the system.
Many healthcare researchers are interested in collecting medical sensor data. As that data may contain many personal facts, many patients are not willing to reveal them.

The sensor network consists of a variety of sensors with a variety of interfaces. A medical provider can set a cloud service, by research organizations, by government-run medical databases, or by private companies providing additional services. The fog layer is the bridge between the sensor network and the cloud. The e-health gateway has sufficient computing capabilities for simple data processing, but its primary role is data aggregation and communication. Depending on the path to the cloud instances, some nodes carry data that has been filtered to enhance privacy. In contrast, other nodes carry data with more physiological data, which might be needed for health monitoring by health care providers or emergency services.

4 Cross4all project model of patient’s healthcare data

The Cross4all project integration model has the patient’s centric concept (Fig. 2). The patients who will have their PHR in the cloud system can store their data and provide privacy and security, but can temporarily grant access to their data to doctors who are also registered in the system. Sensors’ data connected with measuring vital signs of life of patients as healthcare data, are also acquired in order to provide data in their PHR. EHR data also can be taken into consideration by an authorized person without direct connection to PHR by the medical persons or indirectly, from the patient, as scanned unstructured data, also accessible by doctors. Labs and biometrics reports can also be in PHR as scanned unstructured documents optionally.

Medical and omics accessible data can be connected with PHR and related to diseases. Some data can be provided by clinicians and connected with phenotype, metabolic and genetic data and related data with patient’s disease. It means that some recommendations for the patient can be done according to doctors’ insights with a combination of PHR and other available data for the patient. In addition, some soft data related to healthcare data, optionally, can be provided and integrated into PHR, as environmental, social media and other data named as exposome data.

This model can have the potential for healthcare risk assessment for disease taken from PHR and environmental and location connected data. The risk assessment demands the usage of complex algorithms, AI and medical knowledge as well as disease connected data analysis.

Standardization is provided to prevent malicious system misusing. It has to enable security access protocols, intrusion detection and prevention techniques, providing SIEM systems, with audit logs of the users and administrator activities [19].
The integrated model of heterogeneous data into electronic PHR rely on a high level of security and privacy and provide adequate access to data for the appropriate user. In the model, the first step to proper user orientation to the appropriate resource is the Authentication and Authorization sub-system (AAA - Authentication, Authorization and Accounting), for which Keycloak server is used to check the type, credential and the affiliation of user access. The first check for secure access is by verifying the authentication - username and password (which can enter an additional way to verify authentication using a short-term token) to check if the user has the right to access. If it is authenticated, the authorization check is performed, i.e. the role of the user is determined, for example, patient, doctor or pharmacist. The last step in the AAA framework is user accounting, which measures the resources the user consumes during access. This may include the measurement of system time or the amount of data that the user sent and/or received during the session. It also deals with statistical data for sessions and resources using information; it is used for authorization control, billing, trend analysis, resources usage and capacity planning activities.

Fig. 2. Patient’s Electronic PHR – a model of integrated electronic health record.

After this first level of security control, the user is redirected to the appropriate control server in the appropriate domain (here according to the country of origin or affiliation). Distribution should be transparent to users, i.e. the system should have only one unique and integral location for the API URL to be used
by applications and end-user integrations, regardless of the origin of the request. This configuration is possible with two or more servers located in each country, connected to the health-data integration hub (it could be the World Health Organization, for instance) and end-user authentication and authorization requests. Applications are filtered and processed according to the domain of origin of their username, so it is redirected to the appropriate Keycloak server from each country for further processing. Upon completion of the authentication and authorization procedure, the client receives an authenticated token that can be used to access the API endpoints and through them access the EHR data.

Because user access data is disaggregated based on affiliation, specifically on the user’s country of origin, this user identification and authorization data is stored on the federal (shared) server in the respective country and is used for authentication and authorization purpose. The user can be assigned the appropriate role: patient (the most important of which are the role of the patient who owns the PHR data), the role of the physician who can access and generate additional PHR data, and the role of the pharmacist who can access only parts of PHR data related to e-prescription services. Role-based access control for accessing PHR data (or only part of PHR data) is defined in user roles. They are also defined in the Keycloak SSO servers and thus the user gets an authenticated token, which he uses, in the further process. Subsystems that allow routing/redirection to appropriate API endpoints follow these rules, check the authorization token, and grant or deny access to the required data. With this approach, authentication rules can be changed even when the system is in production, and additional segmentation rules for data access can be implemented. Fig. 3 presents Visual Notation for OWL Ontologies (VOWL) used in the model, according to Fast Healthcare Interoperability Resources (FHIR), as a part of PHR software, which integrated EHR.

![Visual Notation for OWL Ontologies (VOWL)](image_url)

**Fig. 3.** Part of the ontology used in the model according to (FHIR).
5 Conclusion

The pandemic situation and the increasing number of patients with chronic diseases demand quick access to patient’s healthcare and medical data. Many problems appear from the lack of healthcare patient’s information, especially when patients with chronic diseases and their treatments are considered. This is very important when patients change the place of living and the medical personnel do not have their data available. The efforts for healthcare data integration until now are mostly intended for high-level decision-makers and data are depersonalized. For this reason, the model is PHR-centric, which integrates data from the PHR of the patient according to HL7 standard. HL7 has developed the FHIR as a new foundation to achieve interoperability. The concept of this model takes into account security and privacy issues, specific for PHR, health and medical-related data and personal data [29]. The model can integrate sensors’ data, exposome and omics data, intended for accurate healthcare risk assessment of the patient, using many public environmental data, atmospheric electromagnetic fields data, social media and other valuable data.

In this model, the patient is the main actor in the system. The patient who has their PHR can temporarily grant access to their data to the selected medical staff and can have the possibility to use PHR and mobile applications to gather healthcare data in their PHR. The medical staff can use mobile applications connected to the specific measurement sensors for professionals to provide a vital signs measurement for the particular patient, acquired and saved in the patient’s PHR securely and privately, according to the country’s data protection law. In the model, PHR is related to data gained from specific medical devices and sensors, from the patient’s EHR as scanned unstructured data, omics data connected with patient’s chronic diseases and social media data. These integrated data have to provide healthcare risk assessment connected with exposome data from environmental databases, connected with the living location of the patient. The implementation of such complex project pointed that such model demands high level of digital healthcare literacy and competency of the patients. As a direction for further works, the model should be validated with exposome data and some algorithms for risk assessment have to be assessed for particular chronic diseases.

6 Acknowledgement

Part of the work presented in this paper has been carried out in the framework of the project “Cross-border initiative for integrated health and social services promoting safe ageing, early prevention and independent living for all (Cross4all)”, which is implemented in the context of the INTERREG IPA Cross Border Cooperation Programme CCI 2014 TC 16 I5CB 009 and co-funded by the European Union and national funds of the participating countries.
References


22. Barouki R, Audouze K, Cournoul X, Demenais F, Gauguier D, Integration of the human exposome with the human genome to advance medicine. Biochimie, Elsevier, 2018, 152, pp.155-158. 10.1016/j.biochii.2018.06.023. hal-02196327


