# Integration of Heterogeneous Data into Electronic Patient Records

Snezana Savoska[*], Blagoj Ristevski[*], Natasha Blazheska-Tabakovska[*], Ilija Jolevski[*], Andrijana Bocevska[*], Vladimir Trajkovik [**]

[*] Faculty of Information and Communication Technologies – Bitola, University “St. Kliment Ohridski” – Bitola, North Macedonia

[**] Faculty of Computer Science and Engineering, “Ss. Cyril and Methodius” University in Skopje North Macedonia

{ snezana.savoska, blagoj.ristevski, natasa.tabakovska, ilija.jolevski, andrijana.bocevska }@uklo.edu.mk, vladimir.trajkovik@finki.ukim.mk

**Abstract - The deficiency of data for patients with chronic diseases and other diseases and previous medical treatments shows a significant weakness with many patients. Typically, due to the healthcare system insufficiently, patients with comorbidities might not survive the diseases, especially when the disease is novel. The lack of information on genetic disorders in patients they are not aware of also contributes to increased patient deaths. This leads to the need to integrate medical and health data with various biological and other data, especially in pandemic circumstances and the increasing number of patients with chronic diseases. Patients' health data issues are evident, but they are stored in various hospital and public health systems such as electronic health records (EHRs), healthcare institutions, and laboratories. Furthermore, biological data are often not integrated and cannot be used by patients, physicians, and specialists to treat particular diseases. Although the urgent need for healthcare and medical data integration is apparent, personal data protection laws are rigorous. They do not allow much progress in the field without implementing patient healthcare data security and privacy standards. One solution for this issue is establishing a personal health record (PHR) as an integrative system for the patient. Many ontological frameworks have been proposed to unify the record formats, but none of them is accepted as healthcare standards. The efforts towards approving the HL7 standards and the well-known medical codding systems promise future data integrations. Also, some attempts are made to associate particular diseases with data obtained from external environmental sensors that measure disease-related data. Using these data, called exposure data or exposome, one can clarify the increasing symptoms of particular diseases influenced by external factors. This paper proposes a cloud-based model for integrating healthcare and medical data from different sources as EHR, health information systems, and measurement sensors into PHR as the first stage towards integrating patient health data. The medical data, PHR, numerous biological and exposome data, and data obtained from wearables are considered and stored on the cloud following the required data security and privacy standards.**

**Keywords- Electronic Health Records, Personal Health Records, Internet of Medical Things, Data Integration.**

## I. INTRODUCTION

The streams of healthcare and medical data nowadays are stored in diverse hospital information systems (HIS) providing protected data for patients in the form of EHR, medical prescription, diseases diagnoses and treatments. These data are owned by many healthcare providers and usually are not accessible to the patients. Taking into account the trends of the population ageing as well as the need for medical care that increases all the time the patients use some wearables according to the concept of IoT and its part Internet of Medical Things (IoMT). They use these wearables for health purposes, according to the concept of Ambient Assisted Living (AAL) that produce many different types of data for the patients’ live conditions and some health parameters. These wearables (sensors) usually are connected with the measurement of some healthcare parameters and collect data for human behavior and health parameters and conditions. If we consider also the exposome concept and all data collected from the environmental pollutions that influence healthcare, we can realize that there are wide amounts of data connected with human health. Recent research shows the increasing interest in the usage of personal patient’s data in the residence where the patient resides at the moment, providing accurate data for healthcare staff that can provide healthcare at the moment when the patient needs medical help and care. The development of artificial intelligence (AI) techniques also can provide a mechanism for healthcare risk assessment when the environmental and chemical exposure data, known as exposome, and other omics and medical data are available. This fact and the usage of this kind of data lead to the need for healthcare and medical data integration.

The proposed solution for this very complex issue is connected with the implementation of strong security standards for healthcare and medical data is the creation of a patient-centric system with Personal Health Record (PHR). This concept considers the responsibility of the patients to secure their PHR and the possibility to share with a selected medical person, temporarily. This concept demands a complex cloud-based architecture as well as the education of the medical staff and patients for increasing digital healthcare literacy. This kind of integrated system also has to support the usage of many wearables connected with mobile applications for measuring and monitoring the vital signs of life, connected with wireless sensors. In addition, the concept has to support input of unstructured data types as healthcare history, medical images, lab results as well as some biological data or data from HIS, EHR, data owned by healthcare providers.

The second challenge nowadays is the patient's healthcare risk assessment. This risk assessment demands the usage of PHR, omics and exposome data [3] that have to be accessible for boots (software agents) for risk assessment. Such complex digital structure is created in a cloud environment, taking into account the security and privacy data protection issues, differ from one country to other. All these different types of collected data, structured, semi-structured and unstructured, have to be saved and acquired according to well-known standards in healthcare as HL7, FHIR, openEHR [16] and to use codding systems as ICD10, standard key terms, such as Medical Subject Headings od SNOMED.

The paper proposes a model of integration of healthcare and medical data in a patient-centric electronic PHR cloud-based system, connected with previously mentioned biological and medical data, omics and exposome data (Fig.1). The purpose is to provide the possibility of combining PHR data with others data types. The paper considers some related works connected with the areas of interest – the attempts for healthcare personal data integration. The third section explains the proposed model of integration of healthcare and medical data and the possibilities for semantic integration of medical, clinical, patient-oriented data, as well as sensors' and exposome data. The final section validates the model and gives some recommendations for future works.

## II. RELATED WORKS

Some many attempts and papers that describe some efforts to integrate healthcare, HIS, clinical and medical data to provide healthcare data analysis intended for healthcare decision-makers. Nevertheless, all these heterogeneous data are stored in many different places, formats and heterogeneous platforms and their integration is a very challenging and demanding task. We can mention Silvestri et al. in [10], in which they proposed a big data architecture for big data analysis considering EHR and PHR data from structured and unstructured HIS document, considering a model for decision makers' support according to the national law, using Spark, MongoDB and DL-bases AI module for NLP. An open data integration platform for patient, clinical, medical and historical data, stored across multiple HISs, is proposed in [13], in which they address patient-centred healthcare and clinical decision support requirements. In this model, they 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 [14]. Serbanati presents a method for digitizing the concept of health by processing the existing information in EHRs with the help of several dedicated services [20]. Precision Medicine includes the discovery of a patient-specific pattern of disease progression and a determination of the precise therapy for that pattern, and the corresponding personalized delivery of care [7]. An 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 [15].

In [8], the authors had demonstrated the feasibility of a scalable, accurate, and efficient approach for medical device surveillance using EHRs, presenting 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. Liang et al. had identified three threats from real cloud-based eHealth systems, i.e., privacy leakage, frequency analysis, and identical data inference [6].

Koren, Jurčević and Huljenić aimed at investigating wireless sensors in the IoT context in contemplation of model solutions in the field of eHealth [8]. Shah and Khan 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 [5]. Gamal, Barakat and Rezk discussed different database models' appropriateness for integrating different EHRs functions with different database specifications and workload scenarios [17]. An application of an unsupervised machine learning approach in discovering latent disease clusters and patient subgroups using EHR data was described in [18]. 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 [19].

Saripalle et al. have described a tethered PHR that seeks to achieve interoperability by using open-source standards and their implementation [9]. Warner and Levy had debated several emerging paradigms for integration including non-standardized efforts between individual institutions and genomic testing laboratories, "middleware" products that portray genomic information, albeit outside of the clinical workflow; and application programming interfaces that have the potential to work within clinical workflow [11].

The conclusion is that the mentioned models indeed try to connect EHR data and provide data integration for decision-makers or some national-wide integration intended to high-level decision-makers in healthcare and medicine. The security issues are considered from the aspect of the patient and the patient's country of living [2]. First, the model is a 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 out of hospitals and HIS and it is not obligatory to be connected with EHR and country of living. This concept demands also an increase in e-health and health digital literacy and can support the national and local medical and healthcare authorities [1] [4]. In addition, data integration should be wider and has to provide data integration not only for data analysis and decision-making but to provide wider integration of a patient-centric electronic PHR system with the possibility to connect them with biological and medical data. The purpose is to provide the possibility of combining PHR data with omics data as well as with open access data from exposome data that affect human health. This concept demands solving the privacy issues of the level of PHRs.

## III. MODEL OF HEALTHCARE DATA INTEGRATION

The model of a PHR-centric integrated healthcare system is built according to HL7 standards with included security and privacy concerns, especially in the part of PHR, e-prescription and the e-referral system [4]. Taking into account that this concept includes the usage of sensors for measuring the vital signs of life, according to the IoT concept, connected with mobile applications for patients and medical practitioners, many security and privacy issues have to be considered. There have to be a couple of scenarios of data integration with the roles of physicians, patients and pharmacists. Patients who have their PHR in the system can have their data stored according to HL7 and FHIR standards and provided privacy and security from cloud systems. The patient temporarily can grant access to their data to medical staff who is also registered in the system. Patients also can use PHR and mobile applications for citizens to collect healthcare data in their PHR.

The doctors can use mobile applications connected with the measurement of a patient’s vital signs of life with sensors for professionals to provide data and to collect them in the PHR. All sensitive data have to be protected and secure in a cloud environment [1]. Users’ data also contain sensitive data such as personal information, health family history, medical and healthcare data, protected properly according to national regulation regarding the personal data protection low in the intended country. Some scanned unstructured data can be entered in the PHR also and they can be accessible to the medical staff. Labs and biometrics reports also can be input into PHR as unstructured documents.

In the next stage, some medical and omics data can be connected with PHR and related to some diseases. These data can be analyzed by clinicians in order to provide some genotype, phenotype and metabolic data related data with some diseases. In addition, some soft data related to healthcare can be provided and integrated into PHR by a patient using environmental, social media and other data known as exposome data.

A suitable structure has to be provided to have available data for healthcare risk assessment for disease taken from PHR and environmental data, connected with location as well as some social media and stress connected data. This extremely complex task has to include data analysis and complex algorithms, AI and medical knowledge as well as risk factor analysis.

When the integration of healthcare and medical data has to be considered, security issues are very important. Security issues can be classified into two main categories: information security and system security. Information security usually means data encryption, data integration and authentication. System security is connected with some administrative, technical and physical security levels. We can use the Authentication, Authorization and Accounting server (AAA) that have to be integrated with attribute authority (providing Role and Routing information). The encryption also has to be used as well as access control to protect and enable PHR security and privacy for the patients’ country of origin. Prevention of unauthorized access to sensitive patient data at rest is provided by data encryption. AAA layer of the system uses the Keycloack Server as an Open Source Identity and Access Management server. Authenticated users (doctors and pharmacists), according to the level of digital identity assurance and patient consent, will have access to this information.

Standardization is provided to prevent malicious system misusing and has to enable security access protocols, intrusion detection and prevention techniques, providing SIEM systems, with audit logs of the users and administrator activities [2]. The proposed model is presented in Fig.1. This model relies on high-level security and privacy and provides adequate access to data for the appropriate user. In the proposed model, the first step to proper user orientation to the appropriate resource is the AAA. It uses the Keycloak server to check type, credential and the affiliation of user access. The first check for secure access is verifying the authentication - username and password to check if the user has the right to access. If the user (patient or doctor) is authenticated, the authorization check is performed, the role of the user is determined. The last step in the AAA framework is user accounting that measures the resources the user consumes during access.

In this first level of security control, the user is redirected to the appropriate control server in the appropriate domain in the country of origin or affiliation. Distribution is transparent to users, only one unique and integral location for the API URL is used by applications and end-user integrations, regardless of the origin of the request. 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. The user can be assigned the appropriate role: a patient who owns the PHR data, the physician who can access and generate additional PHR data, and the pharmacist who can access only parts for e-prescription services. Role-based access control for accessing some part or while PHR data is defined in user roles. 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.

Biological, medical and PHR data usually are stored in relational and non-relational databases containing structured and unstructured parts based on documents that are encrypted at rest. Encrypted data can be further segmented into two parts: a user-identifiable part and a depersonalized part with medical data. Both parts use different encryption keys and thus provide even better security segmentation. This approach provides secure access to personal data from PHR, for specific roles such as a doctor, who need access to data for data analysis. Keycloak allows these privileges to be defined in the role of a doctor, but also allows the identified user to request third-party authorization to access his or her data. The special part is data taken from different owners of data, EHR of the patients, HIS systems, labs, biological data, the data from different bioinformatics databases of omics data related to the specifics of each individual that require special access by the authorized person. They can be imported from physical paper documents in an unstructured format, highly secured and there is no integration of these data with the PHR.

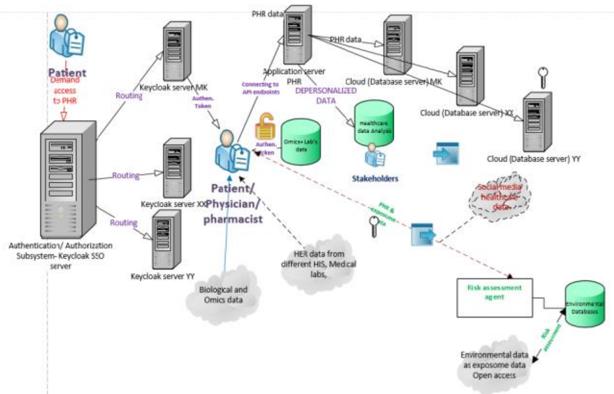


Figure. 1 Model of integration of healthcare heterogeneous data into PHR

The model contains another important part - patient healthcare medical analyzes and assessment of the health risk. The risk can be assessed using the exposome data related to the location where the patients reside and work. These data are usually public and connected with the environmental conditions, saved in time-series data stores, as well as social conditions of living. Data are usually stored in external databases with high scalability and can be used with efficient algorithms for the detection of pollutants and alert of patients with some chronic disease. Data are stored usually stored in time-series and graph-based data repositories with predefined parameters and values and are free for usage. This part should not be subject to high-security criteria for access and the user authentication is a sufficient condition for access to the data.

Access to the base model of integrated electronic PHR can be described in the next steps: First: the user demands access to the system through Keycloak; Second: Keycloak redirects to an appropriate sub-system with a Keycloak server, located in the user's country of origin; Third step: The corresponding Keycloak executes AAA compliance (log of the access -AAA); Fourth: Keycloak assigns an appropriate role to the user.

*The patient* has access to their electronic PHR and free access to the exposome data. These data can be used for some tools for patient's healthcare risk assessment. The patient's healthcare risk assessment can influence the decision of doctors and patients about the lifestyle, place of living, analysis of the patients' medical condition and the doctors' recommendations. The patient has access to the PHR from the encrypted database and can modify their data but not their medical and healthcare data. The patient can add scans from medical paper documents obtained from HIS, lab and other diagnostic data.

*The selected doctor* has access to the electronic PHR given grand permission by the patient. In addition, the doctor has access to exposome data and consider according to the gained health risk assessment for the patient, decide for the needed treatment, give some recommendations for future patient's behavior, avoiding some risk factors associated with particular patient's disease and health risk. Also, the doctor has access to the patient's PHR data, medical and historical data and reports and was granted to enter new measurements for the patient, results obtained with devices based on sensors those measure vital signs, parameters of biometric data.

The doctor has the role and privilege to write an e-prescription or e-referral. With the e-prescription, the doctor shares that part of PHR with the pharmacist and the patient can take the medicine for which his authorized pharmacist has received an e-prescription. The e-referral authorizes the referral physician to be able to see the shared results by the physician who created the referral for the patient.

The role pharmacist was granted access to e-prescription given to the patient in the same domain in which the doctor writes e-prescription. The pharmacist has to provide the drugs or medication to the patient.

## IV. Conclusion

The emerging healthcare situation and today lifestyle demand quick access to patients' healthcare and medical data. Many problems can arise from a lack of healthcare information, especially for patients with chronic diseases. For this reason, the proposed model of a PHR-centric integrated healthcare system is implemented following HL7 standards. This concept of integration and proposed model can solve the security and privacy issues, especially in the part of PHR, e-prescription and the e-referral system. The PHR-centric model integrates heterogeneous data to provide an implementation of evidence-based medicine for the patient who possesses PHR in the place of temporary residing, accurate healthcare risk assessment taking into consideration its PHR, medical, exposome and omics data. The patients who have their PHR can temporarily grant access to their data to the registered medical staff and have the possibility to use PHR and mobile applications to gather healthcare data in their PHR. The medical staff can use mobile applications connected with measurement sensors for professionals to provide a measurement of a patient's vital signs and to collect them in a patient's PHR, using the concept of IoT, IoMT and sensors or wearables. The sensitive personal data are protected and secured properly. In the proposed model, specific medical and omics data are connected with PHR and related to diseases.

The other important advantage is the possibility for healthcare risk assessment connected with the usage of exposome data from environmental databases [12], social media, and other public data that can be connected with location and used by the software agents for healthcare risk assessment [21]. It can be very useful for personal healthcare risk assessment. All these depersonalized medical and healthcare data can be analyzed using suitable big data analytics tools for medical purposes, using disease data for disease group risk assessment.

## 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

[1] Snezana S. et al. (2020) Cloud Based Personal Health Records Data Exchange in the Age of IoT: The Cross4all Project. In: Dimitrova V., Dimitrovski I. (eds) ICT Innovations 2020. Machine Learning and Applications. ICT Innovations 2020. Communications in Computer and Information Science, vol 1316. Springer, Cham.

[2] Snezana Savoska, Ilija Jolevski, Blagoj Ristevski, Natasha Blazeska-Tabakovska, Andrijana Bocevska, Boro Jakimovski, Ivan Chorbev, Vassilis Kilintzis, Design of Cross Border Healthcare Integrated System and its Privacy and Security Issues, In Proceedings of Computer and Communications Engineering, Workshop on Information Security 2019, 9th Balkan Conference in Informatics, Volume 13, 2/2019, pp.58-64

[3] Savoska, Snezana and Ristevski, Blagoj and Blazheska-Tabakovska, Natasha and Jolevski, Ilija (2019) Towards Integration Exposome Data and Personal Health Records in the Age of IoT. 11th ICT Innovations Conference 2019, 17-19 October, Ohrid, Republic of Macedonia.pp.237-246.

[4] Snezana Savoska, Ilija Jolevski, Architectural Model of e-health PHR to Support the Integrated Cross-border Services, In proceedings of ISGT conference 2018, pp. 42-49, Sofia, 2019.

[5] Shahid Munir Shah , Rizwan Ahmed Khan, Secondary Use of Electronic Health Record: Opportunities and Challenges, arXiv:2001.09479v1 [cs.CY] 26 Jan 2020.

[6] Jinwen Liang et al., Privacy-preserving range query over multi-source electronic health records in public clouds, Journal of Parallel and Distributed Computing, Volume 135, January 2020, Pages 127-139.

[7] Sitapati A. et al, Integrated Precision Medicine: The Role of Electronic Health Records in Delivering Personalized Treatment, Wiley Interdiscip Rev Syst Biol Med. 2017 May ; 9(3).

[8] A. Koren, M. Jurčević and D. Huljenić, "Requirements and Challenges in Integration of Aggregated Personal Health Data for Inclusion into Formal Electronic Health Records (EHR),"2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET), Split, Croatia, 2019, pp. 1-5.

[9] Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform. 2019 Jun;94:103188.

[10] Silvestri S. et al., A Big Data Architecture for the Extraction and Analysis of EHR Data, 978-1-7281-3851-0/19/

[11] Warner J. L., Jain S.K., Levy M.A., Integrating cancer genomic data into electronic health records, Warner et al. Genome Medicine (2016) 8:113.

[12] Barouki R. and all, Integration of human exposome with the human genome to advance medicine, Biochimie 152 (2018) pp.155-158.

[13] Jayaratne, M., et al.. A data integration platform for patient-centered e-healthcare and clinical decision support. Future Generation Computer Systems, 92, 996-1008.

[14] Goodrum, H., Roberts, K., & Bernstam, E. V. (2020). Automatic classification of scanned electronic health record documents. International Journal of Medical Informatics, 144, 104302.

[15] S Rubí, J. N., & L Gondim, P. R. (2019). Iomt platform for pervasive healthcare data aggregation, processing, and sharing based on onem2m and openehr. Sensors, 19(19), 4283.

[16] Ristevski, B., & Savoska, S., Healthcare and medical Big Data analytics. Applications of Big Data in Healthcare: Theory and Practice, 85, 2021.

[17] Gamal, A., Barakat, S., & Rezk, A. (2020). Standardized Electronic Health Record Data Modeling and Persistence: A Comparative Review. Journal of biomedical informatics, 103670.

[18] Wang, Y., Zhao, Y., Therneau, T. M., Atkinson, E. J., Tafti, A. P., Zhang, N., ... & Liu, H. (2020). Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. Journal of biomedical informatics, 102, 103364.

[19] Vidal, M. E., Endris, K. M., Jazashoori, S., Sakor, A., & Rivas, A. (2019). Transforming Heterogeneous Data into Knowledge for Personalized Treatments—A Use Case. Datenbank-Spektrum, 19(2), 95-106.

[20] Serbanati, L. D. (2020). Health digital state and Smart EHR systems. Informatics in Medicine Unlocked, 21, 100494.

[21] Savoska, S., Fdez-Arroyabe, P., Cifra, M. et al. Toward the creation of an ontology for the coupling of atmospheric electricity with biological systems. Int J Biometeorol 65, 31–44 (2021).