

Integration of heterogeneous medical and biological data with electronic personal health records

Integration heterogener medizinischer und biologischer Daten in elektronische Patientenakten

Abstract

The shortage of data for patients with chronic and other diseases and previous medical treatments shows significant weakness in the diagnosis and treatment of patients. Due to the healthcare system insufficiency, patients with comorbidities might not survive the diseases, especially when the disease is novel. The lack of information on patients' genetic disorders, especially when they are unaware of them, also contributes to increased patient deaths. This conveys the necessity to integrate medical and health data with various biological omics and other data, especially in pandemic circumstances. Patients' health data matters are apparent, but they are stored in multiple hospitals and 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 regulations are severe. They do not allow much progress in the area without implementing security and privacy standards for patient healthcare data. One solution for this issue is setting 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 a healthcare standard. The efforts toward approving the Health Level Seven (HL7) standards and the common medical coding systems ensure further data integration. Some efforts are made to associate particular diseases with data obtained from external environmental sensors that measure disease-associated data. Using these data, which are called exposome, the increasing symptoms of particular diseases influenced by external factors can be clarified. This paper suggests a cloud-based model for integrating healthcare and medical data from different sources such as EHRs, health information systems, and measurement sensors into the PHR as the first stage toward integrating patient health data. Besides the patients' personal and clinical data, various omics data should be integrated for improved individualized disease prognosis and treatment of the patients. These data are stored in the cloud following the required data security and privacy standards.

Keywords: electronic health records, personal health records, Internet of Medical Things, data integration, omics data, healthcare information system

Zusammenfassung

Der Mangel an Daten über PatientInnen mit chronischen und anderen Krankheiten und medizinischen Vorbehandlungen zeigt eine erhebliche Schwäche bei der Diagnose und Behandlung vieler PatientInnen auf. Aufgrund der Unzulänglichkeit des Gesundheitssystems kann es sein, dass PatientInnen mit Komorbiditäten die Krankheiten nicht überleben, insbesondere wenn es sich um eine neue Krankheit handelt.

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

1 Faculty of Information and Communication Technologies, University "St. Kliment Ohridski" – Bitola, Republic of Macedonia

2 Faculty of Computer Science and Engineering, "Ss. Cyril and Methodius" University of Skopje, Republic of Macedonia

Der Mangel an Informationen über die genetischen Störungen der PatientInnen, vor allem wenn sie sich derer nicht bewusst sind, trägt ebenfalls zu einer erhöhten PatientInnensterblichkeit bei. Daraus ergibt sich die Notwendigkeit, medizinische und gesundheitliche Daten mit verschiedenen biologischen Omics und anderen Daten zu integrieren, insbesondere unter Pandemiebedingungen. Die Relevanz des Themas der Gesundheitsdaten von PatientInnen ist offensichtlich, aber die Daten werden in verschiedenen Krankenhäusern und Gesundheitssystemen wie der elektronischen Patientenakte (ePA), Gesundheitseinrichtungen und Laboren gespeichert. Darüber hinaus werden biologische Daten oft nicht integriert und können von PatientInnen, ÄrztInnen und SpezialistInnen nicht zur Behandlung bestimmter Krankheiten genutzt werden. Obwohl der dringende Bedarf an der Integration von Gesundheits- und medizinischen Daten offensichtlich ist, sind die Vorschriften zum Schutz personenbezogener Daten streng. Sie lassen keine großen Fortschritte in diesem Bereich zu, ohne dass Sicherheits- und Datenschutzstandards für Gesundheitsdaten von PatientInnen eingeführt werden. Eine Lösung für dieses Problem ist die Einrichtung eines Personal Health Records (PHR) als integratives System für die PatientInnen. Viele ontologische Rahmenwerke wurden vorgeschlagen, um die Datensatzformate zu vereinheitlichen, aber keines von ihnen ist als Standard im Gesundheitswesen anerkannt. Die Bemühungen um die Annahme der Health Level Seven (HL7)-Standards und der gängigen medizinischen Codierungssysteme sorgen für eine weitere Datenintegration. Es gibt Bestrebungen, bestimmte Krankheiten mit Daten in Verbindung zu bringen, die von externen Umweltsensoren gewonnen werden, die krankheitsassoziierte Daten messen. Anhand dieser Daten, die als Exposom bezeichnet werden, können die zunehmenden Symptome bestimmter Krankheiten, die durch externe Faktoren beeinflusst werden, geklärt werden. In diesem Artikel wird ein Cloud-basiertes Modell zur Integration von Gesundheits- und medizinischen Daten aus verschiedenen Quellen wie der ePA, Gesundheitsinformationssystemen und Messsensoren in den PHR als erster Schritt zur Integration von Gesundheitsdaten vorgeschlagen. Neben den persönlichen und klinischen Daten der PatientInnen sollen auch verschiedene Omics-Daten integriert werden, um eine bessere individualisierte Krankheitsprognose und Behandlung der PatientInnen zu ermöglichen. Diese Daten werden in der Cloud unter Einhaltung der erforderlichen Datensicherheits- und Datenschutzstandards gespeichert.

Schlüsselwörter: elektronische Patientenakten, persönliche Patientenakten, Internet der medizinischen Dinge, Datenintegration, Omics-Daten, Gesundheitsinformationssystem

I Introduction

Nowadays, healthcare and medical data are stored in diverse hospital information systems providing protected data for patients in the form of electronic health records (EHRs), medical prescriptions, the diagnoses of diseases, and treatments. These data are owned (on disposal) by many healthcare providers and are usually not available to patients. Considering the population ageing trends [1], [2], and the increase of the necessity for medical care, wearables' data that support the ambient assisted living (AAL) concepts enhance the available data. The concepts of the Internet of Things (IoT) and the Internet of Medical Things (IoMT) provide many sensors connected with healthcare that gather data for human behavior and health conditions [3]. Considering the exposome concept

and all environmental data that affect healthcare, it can be apprehended that there are large amounts of data connected to human health [4], [5], [6]. Recent research shows the increasing interest in the use of personal patient data in the residence where the patient stays, providing accurate data for medical staff that should give suitable healthcare when the patient is in need of medical help and care. The development of artificial intelligence (AI) techniques provides a mechanism for healthcare risk assessment when environmental, chemical, exposome, and other omics and medical data are available [7]. All these data need to be integrated into healthcare and medical data.

One of the solutions for this complex issue that is associated with the implementation of strict security standards for healthcare and medical data and which considers the

implementations of GDPR in different countries and Health Level Seven (HL7) as the universal standard is the creation of a patient-centric system where the patient is the data owner of the personal health record (PHR). In this concept, it is the responsibility of the patients to secure their PHR and share it tentatively with selected medical staff. This concept demands a complex cloud-based architecture allowing for patients from different countries and for cross-border travelling for medical purposes as described in [8], [9], [10]. Such an integrated system must support the use of many wearables connected with mobile applications for measuring and monitoring vital signs by using wireless sensors. This concept must also support the input of unstructured data, such as healthcare history, medical images, and laboratory results, as well as some biological data or data from health information systems (HIS), EHRs, and other data owned by different healthcare providers [11]. The patient's healthcare risk assessment is another challenge, connected with the usage of PHR, omics, and exposome data that have to be accessible for risk assessment agents [4], [12]. This complex digital structure has to be based on cloud infrastructure solving the security and privacy data protection issues in different participating countries. Also, it is important to protect healthcare data according to the national legislation as well as to ensure the safety of patients' data by taking into consideration the differences in national data protection laws [9], [13]. The different types of collected data (structured, semi-structured, and unstructured) have to follow well-known standards like HL7, Fast Healthcare Interoperability Resources (FHIR), and openEHR [14], and use coding systems, such as the International Classification of Diseases (ICD-10), and standard key terms, such as Medical Subject Headings (MeSH). These standards and coding systems were accepted by the project's technical committee [15]. More detail can be seen in [10], [11], [15], [16]. Besides patient personal data, such as behavioral and demographic data, clinical data, such as examination, laboratory, and imaging data, as well as omics data (e.g. genomics, proteomics, metabolomics, transcriptomics, lipidomics, epigenomics, microbiomics, immunomics, and exposomics data) should be integrated for enhanced individualized disease prognosis and the treatment of the patients [5], [14].

The proposed model in this paper attempts to prepare the concept of integrating a patient-centric electronic PHR system connected with previously mentioned biological and medical data to provide the possibility of combining PHR data with open-access data from exposome data. The first part of the system was implemented as a project activity in two participating countries (Macedonia and Greece) as a pilot project with the described architecture considering the security and privacy problems in the two participating countries [11], [17].

The rest of the paper is organized as follows. The second section provides related works connected to the fields of interest for healthcare personal data integration, data sources, interoperability of HIS, EHR, electronic patient

records, and other biological, medical, healthcare, and omics data. In section three, the proposed integrative model is depicted and the possibilities for semantic integration of medical, clinical, and patient-oriented data as well as sensors' and exposome data are explained. The advanced possibilities resulting from the integration of omics and exposome data are described in section four. The final section provides the concluding remarks and some directions for future works.

II Related work

Many efforts are made to integrate healthcare, HIS, and clinical and medical data and to enable healthcare data analysis to be suitable for healthcare decision-makers. These heterogeneous data are stored in many different sites, formats, and heterogeneous platforms, and their combination is a very challenging task. For example, Silvestri et al. propose a big data architecture for big data analysis regarding EHR and PHR data from structured and unstructured HIS documents [18]. The interoperability among EHR systems, the full integration of clinical data within PHRs, and the exploitation of the contained information is a widespread target globally.

An open data integration platform for patient, clinical, medical, and historical data spread across various HISs is proposed in [19]. The platform was adopted and implemented to address patient-centered healthcare and clinical decision-support requirements. It can support and integrate further heterogeneous data sources, such as data streams generated by wearable IoT devices. The centralization of data assets permits every stakeholder in a patient-centered care setting to participate in decision-making actively. A wide variety of scanned documents are usually integrated into EHRs. The distribution of scanned documents in one health institution and the design and evaluation of a system to classify documents into clinically relevant and non-clinically relevant categories as well as further sub-classifications are described in [20]. To determine whether this approach could accurately classify scanned documents within an EHR, optical character recognition (OCR) and text classification models trained on documents that were previously manually classified were evaluated. In addition, Serbanati presents a method for digitizing the concept of healthcare by processing the actual information in EHRs with the help of several dedicated services [21].

Precision medicine includes the discovery of a patient-specific pattern of disease progression and a determination of the precise therapy for that pattern, as well as the corresponding personalized care delivery [22]. Although EHRs are instrumental across this spectrum, they focus on personalized healthcare delivery based on advances in genomic medicine. An IoMT platform for pervasive healthcare that ensures interoperability, quality of the detection process, and scalability in a machine-to-machine-based architecture was proposed in [23]. That platform provides functionalities for the processing of

high volumes of data, knowledge extraction, and common healthcare services. In [24], the authors describe the feasibility of a scalable, accurate, and efficient approach to medical device surveillance using EHRs. They describe how implant and implant-related complications as well as notices of post-implant pain can be reliably identified from clinical notes in the EHR. Their methods require orders of magnitude less than labeled training data do and obtain advanced performance.

Liang et al. identify three threats from real cloud-based eHealth systems, i.e. privacy leakage, frequency analysis, and identical data inference [25].

An EHR aggregator (EHRagg) [26] solves the interoperability and accessibility issues using a pragmatic approach by proposing a translation between standards. Wireless sensors in the IoT context in contemplation of model solutions in the field of eHealth focused on merging the PHR collected through wearable and non-wearable sensors into the formal infrastructure, and services within the national central HIS are presented in [24]. Shah and Khan outlined several secondary uses of EHRs 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. Different ethical and privacy issues arising from EHR data leaks are detailed in [27]. Data security and patients' privacy issues related to the secondary uses of EHR, especially when EHR data are transmitted through a network and shared and exchanged with multiple stakeholders, are also studied critically. Gamal et al. discuss the appropriateness of different database models for integrating different EHRs functions with different database specifications and workload scenarios. According to their analysis, every database technology offers diverse healthcare task performance [28].

An application of an unsupervised machine learning approach in discovering latent disease clusters and patient subgroups using EHR data is described in [29]. A knowledge-driven framework able to transform diverse 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 [30]. They describe the application of the framework in the biomedical domain and show the potential to uncover patterns that can explain treatment interactions and patient characterization.

Saripalle et al. present a tethered PHR that achieves interoperability by using open-source standards and their implementation [31]. 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. Warner and Levy discuss several emerging paradigms for integration, including non-stan-

dardized 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 [32]. Zoppi et al. have developed MiBiOmics, a web-based and standalone application that aims to provide established and novel techniques to uncover robust signatures in high-dimensional datasets. This application enables the exploration and integration as well as multi-omics analyses of the detection and description of associations across omics layers and visualization of up to three omics datasets [33].

Xie et al. have developed the Multi-Omics Breast Cancer Database (MOBCdb) as a repository that integrates genomics, transcriptomics, epigenomics, and data from clinical and drug response studies of diverse breast cancer subtypes [34].

All mentioned models attempt to connect EHR data and provide data integration for decision-makers or some national-wide integration intended for high-level decision-makers in healthcare and medicine. They do not consider the patient's data integration.

The model proposed in this paper is a cloud-based cross-border healthcare system based on the PHR concept with an e-health strategy, which was implemented within the project entitled Cross4all [8]. The central point of healthcare data integration is the patient's PHR. The key point is that data acquisition can be made in hospitals and HIS, and it cannot be connected with the EHR and country of living. This concept requires an increase in e-health and digital health literacy, as well as the support of the national and local medical as well as healthcare authorities [8], [15], [35]. Moreover, data integration should be more extensive and has to provide broader data integration, not only for data analysis and healthcare decision-making. In [4] and [8], the integration of healthcare, medical, omics, and sensors data as well as exposure data to provide data for predicting the influence of environmental, social, and stress factors on health was proposed. The proposed model endeavors to prepare the concept of integrating a patient-centric electronic PHR system to connect it with previously mentioned biological and medical data to provide the possibility of integrating PHR data with open access data from exposome data as environmental data from measurements of parameters that affect human health [12]. Security issues in the proposed model are considered from the aspect of the patient and the patient's country of living [16].

III Proposed model of healthcare and medical data integration

The model of a PHR-centric integrated healthcare system is built according to the HL7 standards, which were accepted by the project's technical committee, after detailed study of the project's demands and aims [15]. In addition, the concept includes security and privacy concerns, espe-

cially regarding PHR, e-prescription, and the e-referral system, as it is implemented in the Cross4all project [8], [16], [17]. Considering that such a system can have many sensors associated with mobile applications for patients and medical practitioners, we have to highlight the security and privacy issues connected with these data collections.

III.1 Data integration

There are several scenarios of data integration with the roles of physicians, medical staff, caregivers, patients, and pharmacists [15], [36]. The user with the patient's role, whose PHR is saved in the system, can have his/her data stored according to HL7 and FHIR standards [37] and is provided privacy and security by cloud systems. This user also has the possibility to use their own PHR and mobile applications for citizens to collect healthcare data in their PHR. Also, the user with this role can add his/her PHR scanned unstructured data from medical paper documents obtained from HIS, laboratory, and other diagnostic data. The patient can temporarily grant access privilege to his/her data to medical staff that is also registered in the system [11].

Medical professionals can use mobile applications connected with measurement sensors for professionals to provide a measurement of a patient's vital signs and collect and add them in a patient's PHR. The user with this role can also add laboratory and biometrics reports in PHRs as unstructured documents [3], [38].

Medical and omics data that are associated with the PHR and related to diseases can be provided by clinicians to provide genotype, phenotype, and metabolic data as well as other data associated with some human diseases. Some recommendations for disease diagnosis and patient treatment can be done according to available omics data by suitable specialists. A large amount of data related to healthcare data can be provided and integrated into the PHR by a patient or clinicians using environmental and social media data as well as other data known as exposure data. A suitable structure has to be provided to have available data for healthcare risk assessment for disease taken from the PHR and environmental and location-connected data as well as some social media and stress-associated data. This is an extremely complex task that has to include data analysis, complex algorithms, AI, and medical knowledge as well as risk factor analysis [4], [6], [12].

III.2 Data representation at syntax and semantic level

The PHR data were created according to the available FHIR ontology and the data contained in the ontology [37]. We used one of the widely used standards HL7, and HL7 FHIR is used for exchanging EHRs. In this case, for the Cross4all project needs this standard is used for PHR data as a developer-friendly concept. According to

this standard, an ontology that represents the PHR's domain entities was developed. The ontology was focused on tele-health and integrated care, taking into account medical data as instances of FHIR ontology classes. For this purpose, the methodology for FHIR data types representation was developed in OWL (Web Ontology Language). The backward compatibility with the standard was updated to the latest version of the standard. In the end, we obtain a database according to the publicly available OWL-DL ontology that is given for reuse and future improvement. More detail can be found in [37].

The mentioned ontology was employed to create the basic electronic PHR system for the Cross4all project with the application of the part for tele-health monitoring, integrated health and medical data as well as instances for all types of medical practitioners' investigations. The data integration is embedded on a service layer, with employing of intelligent agents by the PHR owner with other systems and links with other data types and other ontologies' data needed for the integration of the patient data and quantification of the risk for the patient [6] taking into consideration environmental parameters influence.

III.3 Data security

When the integration of healthcare and medical data is considered, security issues are very important.

The PHR contains sensitive user data, such as personal information, health family history, and medical and healthcare data that must be protected and secured.

Security issues can be classified into two main categories: information security and system security. Information security includes data encryption, data integration, and authentication. The sensitive data are protected properly according to national regulations regarding the personal data protection law of the patient's country of living (PHR data are physically residing in the country of patient origin). We used encryption and access control to protect and enable PHR security and privacy. Prevention of unauthorized access to sensitive patient data at rest is provided by data encryption [16], [17].

System security is mostly connected to some administrative, technical, and physical security levels. At this point, we used the authentication, authorization and accounting server (AAA) integrated with attribute authority (providing role and routing information according to the country of living of the patient). The AAA layer of the system is designed to use an open-source identity and access management server (in the project we used Keycloak Server) as authenticated users (medical staff, caregivers, and pharmacists) will have access to these data according to the level of digital identity assurance and patient consent. Standardization also is provided to prevent malicious system misusing. It enables security access protocols, intrusion detection, and prevention techniques, providing SIEM (security information and event management) systems, with audit logs of the users and administrator activities as it was planned in the project [16].

III.4 Proposed model and solutions for security concerns

The proposed model is depicted in Figure 1. The integrated model of heterogeneous data into electronic PHR relies on a high level of security and privacy and provides adequate access to data for the appropriate user. In the proposed model, the first step in matching the proper user orientation to the appropriate resource is the AAA (the Keycloak server is used to check the type, credential, and affiliation of user access). The first check for secure access is by verifying the authentication – username and password (which can be used additionally to verify authentication by 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 as patient, medical staff, 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 is used for authorization control, billing, trend analysis, resource use, and capacity planning activities.

After this first level of security control, the user is redirected to the appropriate control server in the appropriate domain (according to the country of origin or affiliation). Distribution should be transparent for users, i.e. the system should have only a 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 and end-user authentication and authorization requests. Applications are filtered and processed according to the domain of origin of their username, so they are 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 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 role as it is the patient who owns the PHR data), the role of the physician, who can access and generate additional PHR data, and the role of pharmacist, who can access only parts of PHR data related to e-prescription services. Role-based access controls for accessing PHR data (or only part of PHR data) are defined in user roles. They have also been defined in the Keycloak SSO (single sign-on) servers, and thus the user receives an authenticated token which he/she 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 data access segmentation rules can be implemented.

III.5 PHR and external data integration

Biological, medical, and PHR data are stored in relational and non-relational databases containing structured and unstructured parts based on documents that are encrypted at rest. The encryption at rest means that the data staying in the database is strongly encrypted using current encryption algorithms, and the data cannot be accessed even with direct access to the database. 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 in the 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/her data. What is special is that data is taken from different owners: EHRs of the patients, mostly from the HIS systems and laboratories, biological data, and the data from different bioinformatics databases of omics data that are related to the specifics of each individual that require special access by the authorized person.

According to the present concept, the patient does not own EHR data; they can be imported from physical paper documents in an unstructured format. EHR data are highly secured and at this time there is no integration of these data with the patient's electronic PHR. Future integration should be at the level of the patient who possesses that data and can load them into his/her electronic PHR [17]. The model comprises another part that is very important for the patient healthcare and medical analyses and assessment of the health risk: the exposome data related to the location where the patients reside and work. These data are usually public data on the environmental and social conditions of living, stored in external databases with high scalability and with an efficient algorithm for their detection, as well as fast analysis and assessment. They are usually time-series and graph-based data repositories with predefined parameters and values. Using these data in our model assumes prior preparation for use. This part should not be subject to high-security criteria for access, and user authentication is a sufficient condition for access to these data.

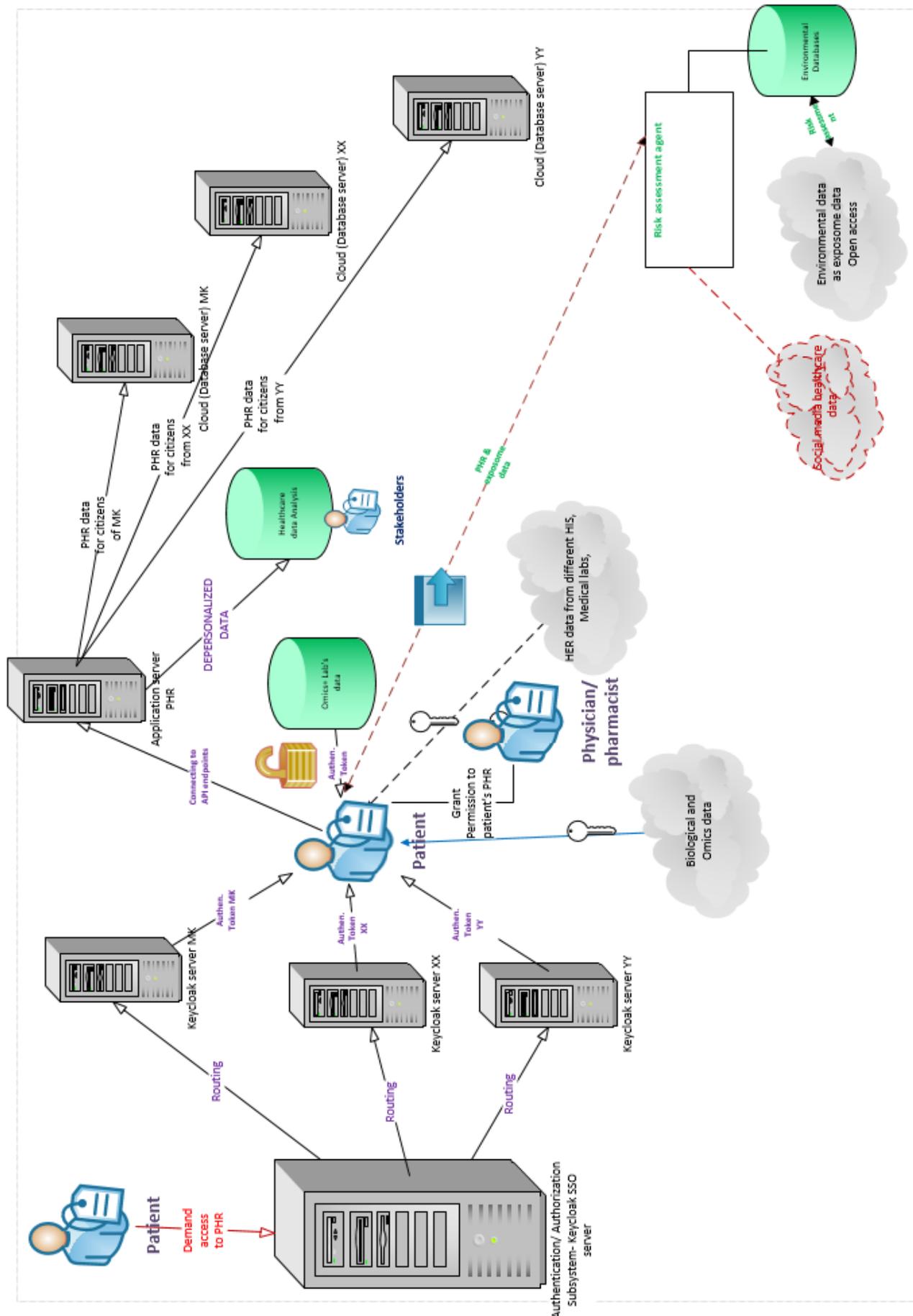


Figure 1: Healthcare security model of electronic PHR integration

III.6 Secure roles in the proposed model

Access to the (base model of) integrated electronic PHR can be described in a few steps. The first four steps are identical for all system users regardless of their roles:

1. The user demands access to the system through Keycloak;
2. Keycloak redirects to an appropriate sub-system with a Keycloak server located in the user's country of origin;
3. the corresponding Keycloak executes AAA compliance (log of the access -AAA);
4. Keycloak assigns an appropriate role to the user. Depending on the assigned role the users have different opportunities [16], [17].

The user with the patient role has access to his/her electronic PHR. This user can also access the exposome data. These data can be used (for some tools) for a patient's healthcare risk assessment depending on the place where the patient resides and works and the disease in his/her PHR, or his/her exposure to some environmental condition for the specific disease, and for offering recommendations. The patient's healthcare risk assessment can be important and influence the decision of doctors and patients about the lifestyle, place of living, analysis of the patient's medical condition, and the recommendations given to them by the doctors. Also, users with this role have access to the PHR from the encrypted database and can modify some of their data but not their medical and healthcare data. These users can add scans from medical paper documents obtained from HIS, or laboratory and other diagnostic data [15].

The user with a medical service provider role (doctors, psychiatrists, dentists, or any other kind of therapist) has access to the electronic PHR given by the patient, but only to the medical analysis and history of medical reports, and was granted to enter new medical or healthcare measurements' data, results obtained with devices based on sensors that measure vital signs, and parameters of biometric data [15]. This user also has access to exposome data and considers these data according to the patient health risk assessment, deciding on the recommendations for the patient's treatment and future patient behavior, and avoiding some risk factors associated with particular environmental conditions of the patient's disease and health risk assessment. Also, the user with this role has the privilege to write an e-prescription for the patient in the same domain as an e-referral [36]. With the e-prescription, the doctor shares that part of the PHR with the pharmacist, and the patient can take the medicine for which his/her 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 [11], [36].

The user with the role of the pharmacist has limited opportunities. He/she has granted access to the e-prescription given to the patient in the same domain in which the

medical staff writes the e-prescription. The pharmacist has to provide the prescribed drugs or medication for the patient.

IV Advanced possibilities for the integration of omics data and exposome data

The integrated PHR owned by the patient is protected and encrypted and can only be accessed by the doctor who has received granted permission from the patient. The module of drugs and medications the patient is prescribed can be accessed by the pharmacist who has been granted the permit by the authorized doctor.

However, when other data that are not related to the PHR database should be included, they are taken at the request of the patient and can have different security preferences. For instance, in the case of omics data, the patient needs to obtain these data from external sources for which he/she needs to have permission to put them on an electronic medium. Analysis of these data by an immunologist or geneticist may indicate a predisposition or presence of a disease associated with the avoidance of certain external conditions of contamination. For instance, if a predisposition for diseases of the airways of the lungs is detected, areas which have a degree of pollution higher than a certain threshold or several stressful situations should be avoided. To obtain the environmental pollution data, the patient uses an intelligent agent, a software module that takes data from public databases that store data on pollutants by location and based on them calculates the pollution level with predefined algorithms and gives recommendations to that patient. These environmental data are usually stored as time-series databases that contain some of the exposome data that are a basis to assess the risk factors for a particular disease according to previously entered assessment factors [4], [12], [39].

V Concluding remarks and further work

The emerging healthcare situation requires quick access to patients' healthcare and medical data, especially during a pandemic. Many problems occur from a lack of data on patients' chronic diseases and the fact that in many countries their treatments are stored in private HISs without the possibility of their integration. For this reason, the proposed model of a PHR-centric integrated healthcare system has to be implemented following HL7 standards [8], [11], [15], [16], [38]. The concept of this model solves the security and privacy issues, especially regarding the PHR, e-prescription, and the e-referral system [36]. Additionally, the model integrates the exposome and omics data intended for an accurate healthcare risk assessment of the patient, taking into consideration

his/her 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 the 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 collect them in a patient's PHR, too. The sensitive personal data must be protected and secured properly according to national regulations regarding personal data protection law in the specific country. In the proposed model, specific medical and omics data are connected with the PHR and related to diseases. The other important issue is healthcare risk assessment connected to the usage of exposome data from environmental databases [4], [5], [6], social media, atmospheric electromagnetic, and other public data that do not have to be protected as PHR but can be connected with location and used by the intelligent agents for healthcare risk assessment by patients and doctors who have the permission to access patient's PHR [8]. It can be very beneficial for personal healthcare risk assessment to know the place of living of a patient [12]. By including omics data and exposure data, the patients can learn a lot about how to take care of themselves and avoid certain activities, choose places to stay, and assess the risk to their health by location, activities, and according to exposure data [39].

This concept can be used for almost any disease. If stress is a factor for a particular disease, wearables may be included. These wearable sensors measure direct or indirect stress factors, calculate stress levels, and make recommendations for avoiding those activities that increase stress in the patient. Smart devices can also control other factors that affect some patients' chronic and other diseases that are detected in their genome. The connection of this data is performed in many parts, and some examples are described in [5].

As a direction for further works, all these depersonalized medical and healthcare data should be stored and analyzed by healthcare and medical authorities using proper big data analytics tools for medical purposes, using disease-associated data for disease group risk assessment [10], [14].

Notes

Authors' ORCIDs

- Prof. Dr. Snezana Savoska:
0000-0002-0539-1771
- Prof. Dr. Blagoj Risteovski:
0000-0002-8356-1203
- Assoc. Prof. Dr. Natasha Blazheska-Tabakovska:
0000-0002-6796-7190
- Prof. Dr. Ilija Jolevski:
0000-0003-2262-9638

- Assoc. Prof. Dr. Andrijana Bocevska:
0000-0001-8701-0700
- Prof. Dr. Vladimir Trajkovik:
0000-0001-8103-8059

Competing interests

The authors declare that they have no competing interests.

References

1. The World Bank DATA. Population ages 65 and above, total. 2010. Available from: <https://data.worldbank.org/indicator/SP.POP.65UP.TO?end=2016&start=1980&view=chart>
2. United Nations, Department of Economic and Social Affairs. World Population Ageing 2015. New York: United Nations; 2015.
3. Savoska S, Blazeska-Tabakovska N, Jolevski I, Bocevska A, Risteovski B, Kilintzis V, Chatzis V, Beredimas N, Maglaveras N, Trajkovik V. Integration of IoMT Sensors' Data from Mobile Applications into Cloud-Based Personal Health Record. In: Antovski L, Armenski G, editors. ICT Innovations 2021. Digital Transformation. 13th International Conference. Revised Selected Papers; 2021 Sep 27-28. Cham: Springer, 2022. (Communications in Computer and Information Science (CCIS); 1521). p. 175-87. DOI: 10.1007/978-3-031-04206-5_13
4. Savoska S, Risteovski B, Blazeska-Tabakovska N, Jolevski I. Towards integration exposome data and personal health records in the age of IoT. In: Gievska S, Madjarov, editors. ICT Innovations 2019. Big Data Processing and Mining. 11th International Conference; 2019 Oct 17-19; Ohrid, North Macedonia. p. 237-46. Available from: <https://proceedings.ictinnovations.org/conferences/2019/17/papers>
5. Barouki R, Audouze K, Coumoul X, Demenais F, Gauguier D. Integration of the human exposome with the human genome to advance medicine. *Biochimie*. 2018 Sep;152:155-8. DOI: 10.1016/j.biochi.2018.06.023
6. Savoska S, Fdez-Arroyabe P, Cifra M, Kourtidis K, Rozanov E, Nicoll K, Dragovic S, Mir LM. Toward the creation of an ontology for the coupling of atmospheric electricity with biological systems. *Int J Biometeorol*. 2021 Jan;65(1):31-44. DOI: 10.1007/s00484-020-02051-3
7. Zdravevski E, Lameski P, Trajkovik V, Pombo N, Garcia N. Importance of Personalized Health-Care Models: A Case Study in Activity Recognition. *Stud Health Technol Inform*. 2018;249:185-8.
8. Savoska S, Jolevski I. Architectural Model of e-health PHR to Support the Integrated Cross-border Services. In: Dimitrov V, Georgiev V, editors. ISGT 2018: Information Systems and Grid Technologies; 2018 Nov 16-17; Sofia, Bulgaria. (CEUR Workshop Proceedings; 2464). p. 42-9. Available from: <https://ceur-ws.org/Vol-2464/>
9. Savoski Z, Savoska S. E-health, Need, Reality or Mith for R. of Macedonia. In: Proceedings of the 8th International Conference on Applied Internet and Information Technologies; 2018 Oct 5; Bitola, North Macedonia. p. 56-9. DOI: 10.20544/AIIT2018.P12
10. Savoska S, Risteovski B, Trajkovik V. Personal Health Record Data-Driven Integration of Heterogeneous Data. In: Dey N, editor. Data-Driven Approach for Bio-medical and Healthcare. Data-Intensive Research. Singapore: Springer; 2023. p. 1-21. DOI: 10.1007/978-981-19-5184-8_1

11. Bocevska A, Savoska S, Risteovski B, Blazheska-Tabakovska N, Jolevski I, Trajkovic V. Cross4all Project Model of Integration of Healthcare Data Using the Concepts of EHR and PHR in the Era of IoT. In: Dimitrov V, Georgiev V, editors. ISGT 2021: Information Systems and Grid Technologies; 2021 May 28-29. (CEUR Workshop Proceedings; 2933). p. 165-77. Available from: <https://ceur-ws.org/Vol-2933/>
12. Radezova Trifunovska M, Jolevski I, Risteovski B, Savoska S. Environmental Data as Exposome and Opportunity of Combining with Cloud-Based Personal Health Records. In: Dimitrov V, Georgiev V, editors. ISGT 2021: Information Systems and Grid Technologies; 2021 May 28-29; Sofia, Bulgaria. (CEUR Workshop Proceedings; 2933). p. 203-19. Available from: <http://ceur-ws.org/Vol-2933/>
13. Assembly of the Republic of North Macedonia. Personal data protection law. RSM Official Journal. 2020 Feb 16;42.
14. Risteovski B, Savoska S. Healthcare and medical Big Data analytics. In: Khanna A, Gupta D, Dey N, editors. Applications of Big Data in Healthcare. London: Academic Press; 2021. p. 85-112. DOI: 10.1016/B978-0-12-820203-6.00005-9
15. Savoska S, Kilintzis V, Jakimovski B, Jolevski I, Beredimas N, Mourouzis A, Corbev I, Chouvarda I, Maglaveras N, Trajkovic V. Cloud Based Personal Health Records Data Exchange in the Age of IoT: The Cross4all Project. In: Dimitrova V, Dimitrovski I, editors. ICT Innovations 2020. Machine Learning and Applications. 12th International Conference. Proceedings; 2020 Sep 24-26; Skopje, North Macedonia. Cham: Springer; 2020. (Communications in Computer and Information Science (CCIS); 1316). p. 28-41. DOI: 10.1007/978-3-030-62098-1_3
16. Savoska S, Jolevski I, Risteovski B, Blazheska-Tabakovska N, Bocevska A, Jakimovski B, Chorbev I, Kilintzis V. Design of cross border healthcare integrated system and its privacy and security issues. Computer and Communications Engineering. 2019;13(2):58-63.
17. Savoska S, Jolevski I, Risteovski B, Blazeska-Tabakovska N, Bocevska A, Kilintzis V, Chorbev I, Jakimovski B. Model Of A Healthcare Integrated System And Solving Security And Privacy Matters In Personal Health Records. In: Kaloyanova K, editor. Information Security in Education and Practice. Newcastle upon Tyne: Cambridge Scholar Publishing; 2021. p. 29-47.
18. Silvestri S, Esposito A, Gargiulo F, Sicuranza M, Ciampi M, De Pietro, G. A big data architecture for the extraction and analysis of EHR data. In: Chang CK, Chen P, Goul M, Oyama K, Reiff-Marganec S, Sun Y, Wang S, Wang Z, editors. 2019 IEEE World Congress on Services (SERVICES); 2019 Jul 8-13; Milan, Italy. New York: IEEE Computer Society Conference Publishing Services (CPS); 2019. p. 283-8. DOI: 10.1109/SERVICES.2019.00082
19. Jayaratne M, Nallaperuma D, De Silva D, Alahakoon D, Devitt B, Webster KE, Chilamkurti N. A data integration platform for patient-centered e-healthcare and clinical decision support. Future Generation Computer Systems. 2019 Mar 1;92:996-1008. DOI: 10.1016/j.future.2018.07.061
20. Goodrum H, Roberts K, Bernstam EV. Automatic classification of scanned electronic health record documents. Int J Med Inform. 2020 Dec;144:104302. DOI: 10.1016/j.ijmedinf.2020.104302
21. Serbanati LD. Health digital state and Smart EHR systems. Informatics in Medicine Unlocked. 2020 Jan 1;21:100494. DOI: 10.1016/j.imu.2020.100494
22. Sitapati A, Kim H, Berkovich B, Marmor R, Singh S, El-Kareh R, Clay B, Ohno-Machado L. Integrated precision medicine: the role of electronic health records in delivering personalized treatment. Wiley Interdiscip Rev Syst Biol Med. 2017 May;9(3):e1378. DOI: 10.1002/wsbm.1378
23. Rubi JNS, Gondim PRL. IoMT Platform for Pervasive Healthcare Data Aggregation, Processing, and Sharing Based on OneM2M and OpenEHR. Sensors (Basel). 2019 Oct;19(19):4283. DOI: 10.3390/s19194283
24. Koren A, Jurčević M, Huljenić D. Requirements and challenges in integration of aggregated personal health data for inclusion into formal electronic health records (EHR). In: Ivšić B, Petrović G, Dadić M, editors. 2019 2nd International Colloquium on Smart Grid Metrology (SMAGRIMET); 2019 Apr 9-12; Split, Croatia. New York: IEEE; 2019. p. 1-5. DOI: 10.23919/SMAGRIMET.2019.8720389
25. Liang J, Qin Z, Xiao S, Zhang J, Yin H, Li K. Privacy-preserving range query over multi-source electronic health records in public clouds. Journal of Parallel and Distributed Computing. 2020 Jan 1;135:127-39. DOI: 10.1016/j.jpdc.2019.08.011
26. Prados-Suárez B, Molina C, Peña Yañez C, Prados de Reyes M. Improving electronic health records retrieval using contexts. Expert Systems with Applications. 2012 Aug 1;39(10):8522-36. DOI: 10.1016/j.eswa.2012.01.016
27. Shah SM, Khan RA. Secondary use of electronic health record: Opportunities and challenges. IEEE Access. 2020 Jul 22;8:136947-65. DOI: 10.1109/ACCESS.2020.3011099
28. Gamal A, Barakat S, Rezk A. Standardized electronic health record data modeling and persistence: A comparative review. J Biomed Inform. 2021 Feb;114:103670. DOI: 10.1016/j.jbi.2020.103670
29. Wang Y, Zhao Y, Therneau TM, Atkinson EJ, Tafti AP, Zhang N, Amin S, Limper AH, Khosla S, Liu H. Unsupervised machine learning for the discovery of latent disease clusters and patient subgroups using electronic health records. J Biomed Inform. 2020 Feb;102:103364. DOI: 10.1016/j.jbi.2019.103364
30. Vidal ME, Endris KM, Jazashoori S, Sakor A, Rivas A. Transforming heterogeneous data into knowledge for personalized treatments – A use case. Datenbank-Spektrum. 2019 Jul 15;19(2):95-106. DOI: 10.1007/s13222-019-00312-z
31. Saripalle R, Runyan C, Russell M. Using HL7 FHIR to achieve interoperability in patient health record. J Biomed Inform. 2019 Jun;94:103188. DOI: 10.1016/j.jbi.2019.103188
32. Warner JL, Jain SK, Levy MA. Integrating cancer genomic data into electronic health records. Genome Med. 2016 Oct;8(1):113. DOI: 10.1186/s13073-016-0371-3
33. Zoppi J, Guillaume JF, Neunlist M, Chaffron S. MiBiOmics: an interactive web application for multi-omics data exploration and integration. BMC Bioinformatics. 2021 Jan;22(1):6. DOI: 10.1186/s12859-020-03921-8
34. Xie B, Yuan Z, Yang Y, Sun Z, Zhou S, Fang X. M0BCdb: a comprehensive database integrating multi-omics data on breast cancer for precision medicine. Breast Cancer Res Treat. 2018 Jun;169(3):625-32. DOI: 10.1007/s10549-018-4708-z
35. Blazheska-Tabakovska N, Risteovski B, Savoska S, Bocevska A. Learning management systems as platforms for increasing the digital and health literacy. In: ICEBT '19: Proceedings of the 2019 3rd International Conference on E-Education, E-Business and E-Technology; 2019 Aug 2-4; Madrid, Spain. New York: Association for Computing Machinery; 2019. p. 33-7. DOI: 10.1145/3355166.3355176
36. Kitanovski D, Stojmenski A, Chorbev I, Jakimovski B, Savoska S, Risteovski B, Jolevski I, Blazheska Tabakovska N, Bocevska A. Implementation of a Cloud-Based Personal Health System for Cross-Border Collaboration. In: Antovski L, Armenski G, editors. ICT Innovations 2021. Digital Transformation. 13th International Conference; 2021 Sep 27-28. p. 17-28. Available from: <https://proceedings.ictinnovations.org/conferences/2021/19/papers>
37. Kilintzis V, Kosvyra A, Beredimas N, Natsiavas P, Maglaveras N, Chouvarda I. A sustainable HL7 FHIR based ontology for PHR data. Annu Int Conf IEEE Eng Med Biol Soc. 2019 Jul;2019:5700-3. DOI: 10.1109/EMBC.2019.8856415

38. Bocevska A, Savoska S, Jolevski I, Blazheska-Tabakovska N, Risteovski B. Implementation of Innovative e-Health Services and Digital Healthcare Ecosystem - Cross4all Project Summary. In: Dimitrov V, Georgiev V, editors. ISGT 2022: Information Systems and Grid Technologies; 2022 May 27-28; Sofia, Bulgaria. (CEUR Workshop Proceedings; 3191). p. 285-301. Available from: <https://ceur-ws.org/Vol-3191/>
39. Savoska S, Bocevska A, Simonoski H. Visualization of Sensors' Data in Time Series Databases for Health Purposes. In: Dimitrov V, Georgiev V, editors. ISGT 2022: Information Systems and Grid Technologies; 2022 May 27-28; Sofia, Bulgaria. (CEUR Workshop Proceedings; 3191). p. 317-34. Available from: <http://ceur-ws.org/Vol-3191/>

Please cite as

Savoska S, Risteovski B, Blazheska-Tabakovska N, Jolevski I, Bocevska A, Trajkovik V. Integration of heterogeneous medical and biological data with electronic personal health records. *GMS Med Inform Biom Epidemiol.* 2023;19:Doc01.
DOI: 10.3205/mibe000240, URN: urn:nbn:de:0183-mibe0002403

This article is freely available from

<https://doi.org/10.3205/mibe000240>

Published: 2023-07-04

Copyright

©2023 Savoska et al. This is an Open Access article distributed under the terms of the Creative Commons Attribution 4.0 License. See license information at <http://creativecommons.org/licenses/by/4.0/>.

Corresponding authors:

Prof. Dr. Snezana Savoska
Faculty of Information and Communication Technologies,
University "St. Kliment Ohridski" – Bitola, Partizanska bb
7000 Bitola, Republic of Macedonia
snezana.savoska@uklo.edu.mk

Prof. Dr. Blagoj Risteovski
Faculty of Information and Communication Technologies,
University "St. Kliment Ohridski" – Bitola, Partizanska bb
7000 Bitola, Republic of Macedonia
blagoj.risteovski@uklo.edu.mk