Decentralized e-Health Architecture for Boosting Healthcare Analytics

Igor Kotsiuba
G. E. Pukhov Institite for Modeling in Energy Engineering, National Academy of Sciences of Ukraine
Kyiv, Ukraine
i.kotsiuba@gmail.com

Artem Velykzhanin
Computer Science and Engineering department, Volodymyr Dahl East Ukrainian National University
Severodonetsk, Ukraine
velykzhanin@snu.edu.ua

Yury Yanovich
Bitfury
Amsterdam, Netherlands
yury.yanovich@bitfury.com

Inna Skarga-Bandurova
Computer Science and Engineering department, Volodymyr Dahl East Ukrainian National University
Severodonetsk, Ukraine
skarga-bandurova@snu.edu.ua

Yuriy Dyachenko
Department of International Economics and Tourism, Volodymyr Dahl East Ukrainian National University
Severodonetsk, Ukraine
yuriy.y.dyachenko@gmail.com

Viacheslav Zhygulin
Bitfury
Amsterdam, Netherlands
viacheslav.zhygulin@bitfury.com

Abstract—In this article, we present an overview of the problems associated with the analysis and security of medical data and offer a solution that will provide the basis for improving the quality of medical services. We propose the architecture of a decentralized health data ecosystem based on a blockchain that will allow us to operate with vast volumes of clinical data, while also protecting confidential medical data. An example of a blockchain solution based on Exonum framework for state-scale use in healthcare is discussed. The deployments of such systems will the benefit to medical data safety, extend the base of clinical data collections, and create an effective shared health infrastructure.

Keywords—electronic health record, analytics, architecture, blockchain

I. INTRODUCTION

The digital transformation of healthcare has made the medical information and services globally accessible. Typically, a patient may have a variety of health care providers, including primary care physicians, specialists and clinicians who generate various patient medical data. These data can be used by different specialists to implement different medical services. However, the use of these disparate data sources requires the right tools and competencies that are not always well developed. It is important to take into account several points:

- Medical data is very extensive and cumbersome.
- There is a problem of the quality of medical data, which complicates the analysis, diagnosis and prognosis.
- Confidentiality is another major public health problem. In this regard, there is a need for additional protection of information in this area, especially in connection with the growing number of cyber crimes.

For healthcare organizations, this can mean the beginning of profound changes in business models aimed at creating a distributed and secure personal data market, and using effective means of medical analytics. Thus, the overall goal is to create a suitable platform for the development of a data infrastructure that relies on general medical data standards such as ICD11, HL7, and provides users with access to analytics generated from real-time data sources, for example directly from electronic health records (EHR).

The paper is organized as follows: section 2 is comprised of overview the healthcare data issues and related work in the scope of healthcare analytics in e-health systems. Afterwards, it is given a description of the proposed architecture. The short overview of blockchain solution for state-scale use in healthcare is given in section 4 followed by the conclusion in section 5 as well as pointing out further developments.

II. HEALTHCARE DATA ISSUES

A. Healthcare data analytics issues

Medical data is one of the most difficult types of data in the study. Health care providers track patients' visits through the EHR system, creating terabytes of medical records. In this case, the EHR includes [1]:

- Electronic records of visits—medical examinations, the results of specialist consultations, nurses records, test results, research results, etc.

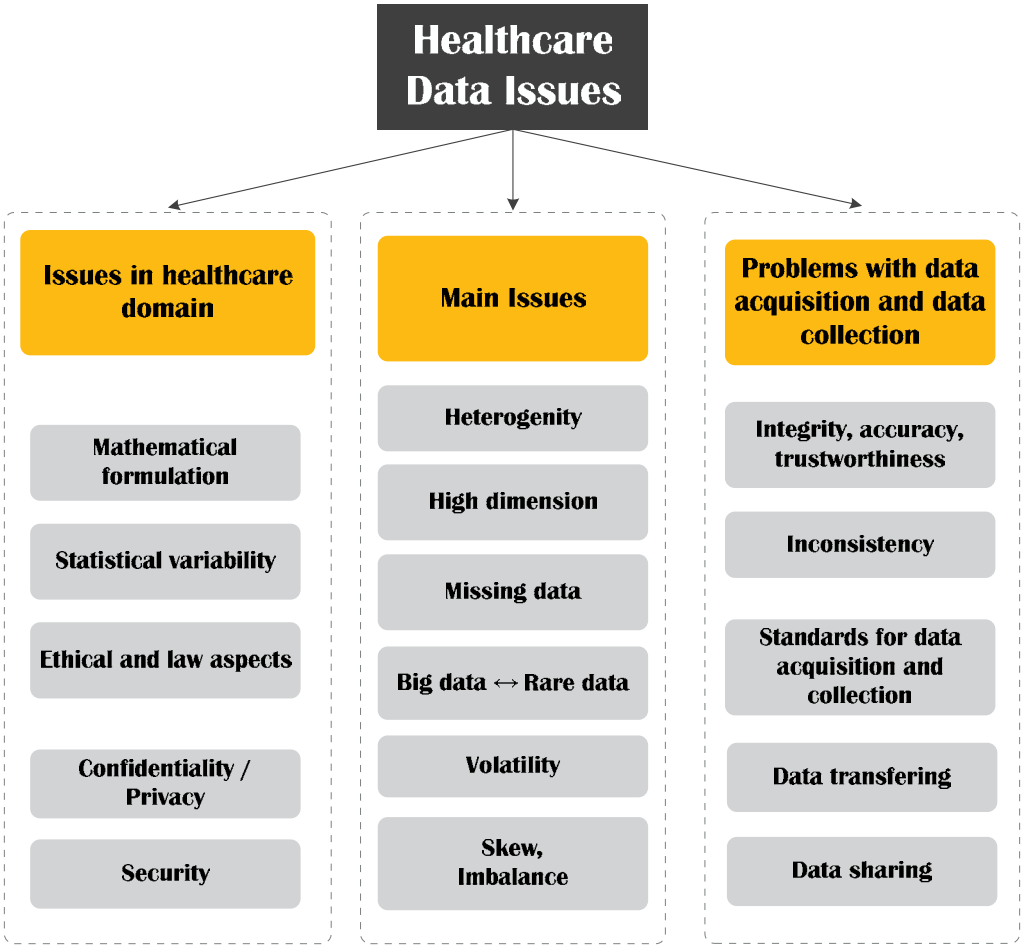

Figure 1. Main problems with healthcare data (Adapted from [3])

- Appointments — all types of appointments to the patient in cases of treatment (laboratory tests, diagnostic manipulations, medications — including preferential prescriptions, specialist advice, etc.)
- Results of laboratory and instrumental studies, scanned and digital medical images.
- Graphic files, scanned images, digital photos, allowing to assess the severity of the course of the disease and the dynamics of the state of the process during the treatment of the patient.

In addition to EHR, medical data may contain other types of data that are an important source for medical analytics:

- Administrative data.
- Claims data.
- Patient/Disease registries.
- Health surveys.
- Clinical trials data.

Some other types of data, such as videos, audio recordings, information about the economic status of patients, can also have significant prognostic significance for medical analytics.

Although the amount of health-related data and the number of medical projects is increasing, effective data analysis is problematic [2]. This is due to a number of reasons, among which the quality of data has long been recognized as a weighty characteristic that affects the quality of the results of data analysis. In the scientific literature, the discussion of the problems of medical data is mainly focused on the purity of data from the point of view of measurement errors, missing values, the presence of incorrect data, and so on. The problems of medical data can be divided into different groups using different criteria. The main problems of medical information are classified and presented in Fig.1.

In a number of works, the main signs of poor-quality data are displacement, imbalance, heterogeneity, and the presence of a large number of missing data. It is difficult to obtain a qualitative analysis result the availability in one set of data of various types, high dimensionality of the data.

For example, [4] distinguishes between two types of data quality problems: incomplete data and incorrect data. In the study [5], problems, which make it difficult to obtain a qualitative result of predicting the conditions of patients in the intensive care unit, are determined by the high dimensionality of the data, their imbalance and,

given that the analysis is carried out in real time, temporary data asynchronization. In [6], too, the state of the patients of resuscitation department is investigated and the main problem of qualitative state prediction is the missing data. A review of leading medical journals by the authors of [7] showed that no data are common in randomized trials with results from patients.

Skewed data is also a common problem in studies with longitudinal, spatial, multilevel, or multidimensional medical data [8]. It is noteworthy that even high-quality medical data, as a rule, are very heterogeneous and complex and require special approaches for preliminary processing and analysis.

In the previous study [9], we developed and tested a general strategy for working with missing data for small sets of patient clinical data, which helps to improve the quality of data analysis, in particular for predictive analytics, but only approximates the desired solution. The problem of using data from centralized local databases is that in a number of cases, for example, for rare diseases, information is not sufficient to make informed decisions.

On the other hand, most large healthcare providers already have the data resources necessary to locate disease markers and implement medical analytics, but accessing this information — and replenishing it with external data — can certainly be a problem. The difficulty is exacerbated when patients move between providers, health systems or geographic regions, as is often the case. The primary health care provider in one region may not receive important information about the patient, because in another region it is unlikely that these systems will be connected.

Thus, it should be noted that there is a clear need for new global integration approaches to health care.

B. Healthcare data privacy

Another important issue of digital health data is to ensure their confidentiality and safety. HIPAA security provisions include a long list of specifications for organizations that store personal health information (PHI), including transmission security, authentication protocols and access control, integrity and audit tools. Nevertheless, to date, healthcare data are subject to an almost infinite number of vulnerabilities.

In addition, to the traditional set of vulnerabilities, new ones are constantly added. So, Internet of Things, the revolutionary technology of the last decades, has brought the technology of medical care for patients with chronic diseases to a new level. At the same time, connecting remote patient monitoring tools, and transferring health data streams to the clinical environment, causes new difficulties associated with real-time information broadcasting. Most people believe that their medical and other medical information is private and must be protected, and patients usually want to know how this information is handled [10].

Currently, patients do not have control over the rights of access to their medical records and are not aware of the real value of the data that they have, which also significantly hampers the development of large analytics.

Therefore, approaches are needed that not only ensure the effective processing and use of data, but also guarantee the protection of confidential data and allow the patient to manage access, and benefit from their data, as a reward contribution to overall medical progress.

III. DECENTRALIZED PLATFORM

As one of the solutions to these problems, we propose the use of a decentralized architecture based on the Exonum open source platform for the healthcare market. Blockchain technology [11] was previously proposed as a storage medium and data sharing in e-health systems [12, 13, 14], to control the accesses to EHR and EMR [15, 16] and as a tool for improving data transparency [17]. Converging blockchain and artificial intelligence technologies is discussed in [18].

The architecture of the proposed platform includes two parts, open and closed (Fig. 2). The system is a distributed storage of medical card data and patient health information. In the closed part, personal medical data is stored, it is a distributed EHR system, the doctor adds it in the record, signing it with his own electronic digital signature (EDS).

At the beginning, the process of data validation takes place, the patient's data are considered valid if they have a digital signature of the doctor. Then the process of data depersonalization occurs, the personal unique identifier of the user is deleted. The data is loaded into the cloud storage (the open part of the system) and in the subsequent can be used by doctors and scientists, for use in large analytics.

Also, open data can have other sources of information, such as diagnostic centers, Internet things, wearable, smart devices.

Services are the links between parts of the system.

In this structure, blockchain technology acts as a mechanism for monitoring and recording data

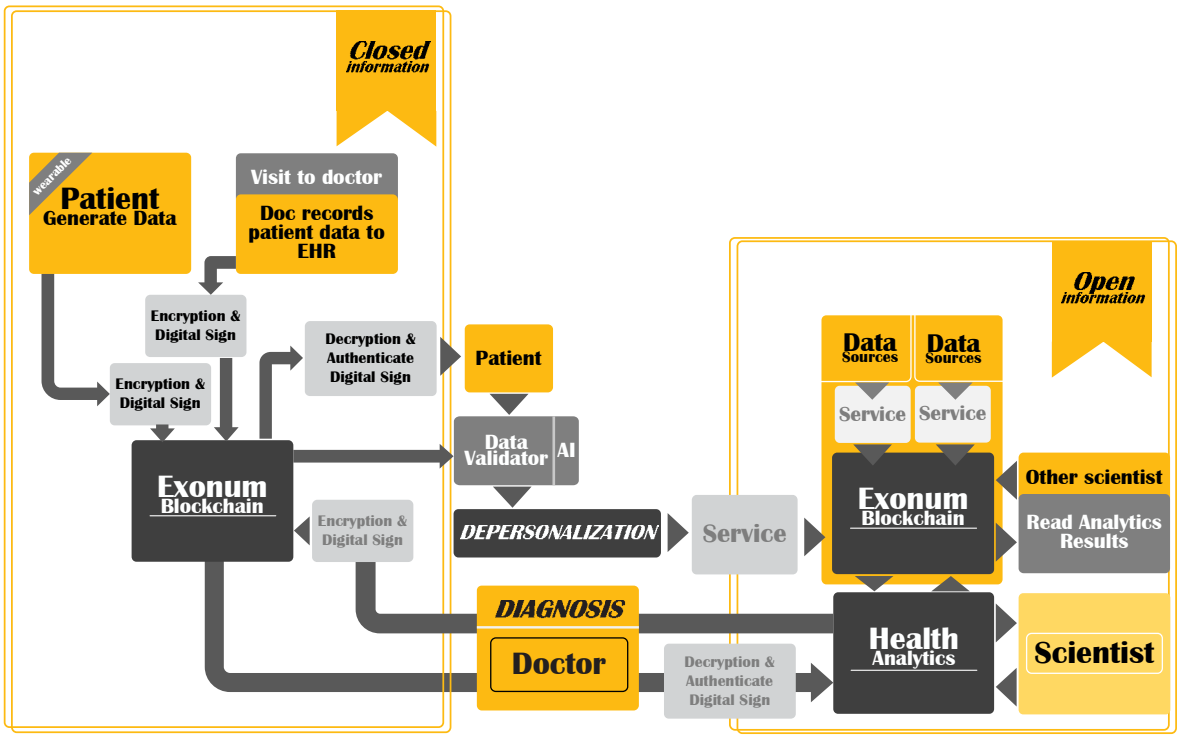

Figure 2. An architecture of e-health platform for healthcare analytics

on changes in medical records. Decentralized architecture involves storing data in several nodes, which can be simply databases or host computers. Data stored in these nodes is usually replicated or, and the technology of distributed ledgers provides quick access to data for this large number of nodes. In terms of data management, blockchains provide the capabilities of a transactional system. Block producer processes transactions that are very similar to transactions in conventional database management systems and accordingly changes the stored values.

This will allow users to create a patient profile to track their medical history and provide access to specialists from different medical organizations on the same platform where they can view the same data for common patients.

The software provides patients with the ability to track visits to doctors, medical bills, personal health information, insurance, immunization and pharmacy products. The user has the opportunity to share their medical information.

Access to depersonalized data will allow doctors (clinicians), managers and researchers to use large data analytics, which means improved early diagnosis technologies and opportunities for better care. On the other hand, by investing in data banks, patients will be able to access personalized analytics, taking into account general indicators.

IV. THE BLOCKCHAIN SOLUTION FOR STATE-SCALE USE IN HEALTHCARE

Blockchain is a continuous sequential chain of blocks (linked list), which can be used to create a distributed register of health data records. The technology of the block chain allows creating open and thus protected systems [19]. After writing data to the block, they can not be changed afterwards without changing all previous blocks and the consent of network members [20]. The proposed solution is built on the basis of Exonum Fig. 3.

Exonum [21] is a framework, not a ready-made block system (like Bitcoin). It consists of complete nodes connected through peer-to-peer connections, and light clients. Full nodes copy all the contents of the block-book and correspond to replicas in the distributed databases. All full nodes are authenticated using a public-key cryptosystem. Complete nodes are divided into 2 categories:

- Auditors.
- Validators.

Auditors have a complete copy of all the information from the blockchain. They can check the whole blockchain consistency, but they can not choose which transactions should be included in the blocks (that is, they can not generate new blocks).

Validators ensure the viability of the network. Only validators can generate new blocks using the

Figure 3. The blockchain solution for state-scale use in healthcare

Byzantine fault-tolerant (BFT) consensus [22, 23]. The validators receive transactions, check them and include them in a new block. The list of validators is limited to network administrators and should consist of 4–15 nodes.

Exonum uses a service-oriented architecture (SOA) [21] and architecturally consists of three parts: services, clients and middleware.

- Services are the main point of extensibility of the structure, which encapsulates the business logic of the blockchain applications.

Like smart contracts on some block platforms, the service defines the transaction rules, the service state, converted by transactions, is stored as part of the overall storage block, the service can also allow external clients to read the relevant data

from the current state of the block. Each service has a well-defined interface to communicate with the outside world, which is essentially a set of endpoints, and the implementation of the interface.

- Light clients realize the typical functionality of customers in SOA.

A light client is a JavaScript library with a number of auxiliary functions available for use by frontend developers. These helper functions are used to test responses from client- side blocking using cryptographic proof. For an easy client, there are two typical uses: the formation and sending of transactions to the Exonum blockchain network, the formation of requests to the full nodes of the network (usually, HTTP GET requests) and verification of responses

- Middleware provides ordering and atomicity of transactions, interaction between services and clients, replication of services between nodes on the network (which are designed for both fault tolerance of service and for audit through audit nodes), service life-cycle management (for example, service deployment), data consistency , access control, help in generating responses to reading requests, etc. That is, the middleware reduces the complexity of the system from the point of view of service developers.

Services will interact with the outside world through 3 types of transactions, read requests and a private API. Transactions are necessarily authenticated by their compilers using digital signatures with a public key to ensure their integrity, as well as in real time and retrospective universal verifiability. The public key infrastructure (PKI) in this case is built from above, in order to provide more complete irrevocable and/or fine-grained access control, if necessary.

Read requests can be processed locally by any complete node (medical institution or organization that has sufficient read access to the corresponding key spaces of block-chain states).

At the same time, the patient gets access by means of an light client, which can communicate with the complete nodes (ie, call the service endpoints and receive answers). A block- based system allows a user to upload their data directly to the system and give their permission to use PHI using a transparent price formula determined by the data model. It also ensures fair tracking of all data usage activities.

At the same time, it becomes possible to link several sources of information together to specialized data banks (Fig. 3), accessed through a query system that can form massive collections of large data and provide a deeper and more effective picture of the history, diagnoses, treatment, socioeconomic problems and patient risk profiles.

V. CONCLUSION

The use of blockchain in the health ecosystem will allow us to operate with huge volumes of clinical data, while also protecting confidential medical data. Another important advantage is the ability to create a data-based market where patients will receive a personal data monitoring tool that will allow them to actively participate in accelerating medical analytics and receive rewards for providing their data to medical institutions, pharmaceutical companies and research institutions.

The development and implementation of a large data analysis program will allow us to move from the big data paradigm to the paradigm of smart data. We assume that the use of such systems will allow profound changes in business models in healthcare and ensure a transition to a new level of provision of medical services.

Certainly, there are a number of unsolved problems, starting with the legal aspects of implementing such systems and ending with the problems of technical compatibility and acceptance by people that need to be resolved in the coming years. At the same time, the very idea of creating massive data collections promises explosive prospects in the diagnosis and treatment of many diseases.

References

[1] P. C. Tang, J. S. Ash, D. W. Bates, J. M. Overhage, and D. Z. Sands, Personal health records: definitions, benefits, and strategies for overcoming barriers to adoption, Journal of the American Medical Informatics Association, 2006, vol. 13, No. 2, P. 121–126.

[2] J. Bresnick, The difference between big data and smart data in healthcare. Health Analytics. Available: https://healthitanalytics.com/features/the-difference-between-big-data-and-smart-data-in-healthcare.

[3] N. Esfandiari, M. R. Babavalian, A. M. E. Moghadam, and V. K. Tabar, Knowledge discovery in medicine : Current issue and future trend. Expert Systems with Applications, 2014, 41(9). P. 4434–4463.

[4] X Wu., V. Kumar, J. R. Quinlan, J. Ghosh, Q. Yang, H. Motoda, G. J. McLachlan, A. Ng, B. Liu, S. Y. Philip and Z. H. Zhou, Top 10 algorithms in data mining. Knowledge and information systems, 2008, 14(1). P. 1–37.

[5] J. Liu, X. X. Chen, L. Fang, J. X. Li, T. Yang, Q. Zhan, K. Tong and Z. Fang, Mortality prediction based on imbalanced high-dimensional ICU big data. Computers in Industry, 2018, 98. P. 218–225.

[6] A. Nagrebetsky and E. A. Bittner, Missing Data and ICU Mortality Prediction: Gone But Not to Be Forgotten. Critical care medicine, 2017, 45(12). P. 2108–2109.

[7] D. Scharfstein Final Report : Sensitivity Analysis Tools for Randomized Trials with Missing Data, 2017. p. 112.

[8] B. M. Ringham, S. M. Kreidler, K. E. Muller and D. H. Glueck, On the distribution of summary statistics for missing data. Communications in Statistics-Theory and Methods, 2018. P. 1–17.

[9] I. Skarga-Bandurova, T. Biloborodova, Y. Dyachenko, Strategy to Managing Mixed Datasets with Missing Items. International Conference on Information Processing and Management of Uncertainty in Knowledge-Based Systems. Springer, Cham, 2018. P. 608–620.

[10] Office for Civil. Your Rights Under HIPAA. HHS.gov. US Department of Health and Human Services. 2017. Available: https://www.hhs.gov/hipaa/for-individuals/guidance-materials-for-consumers/index.html.

[11] S. Nakamoto, Bitcoin: A Peer-to-Peer Electronic Cash System. 2008. www.Bitcoin.org, 9. https://doi.org/10.1007/s10838-008-9062-0.