Blockchain Based Network for Tuberculosis: A Data Sharing Initiative in Brazil

Filipe BERNARDI[a,1], Vinicius LIMA[a], Felipe PELLISON[a], Paulo Mazzoncini de Azevedo MARQUES[a,d], Rui RIJO[b,e,f], Rafael Mello GALLIEZ[c], Afrânio KRITSKI[g] and Domingos ALVES[d,e]

[a] *Bioengineering Postgraduate Program, University of São Paulo, Brazil*
[b] *School of Technology and Management, Polytechnic Institute of Leiria, Portugal*
[c] *Center of Scientific and Technological Information, Oswaldo Cruz Foundation, Brazil*
[d] *Ribeirão Preto Medical School, University of São Paulo, Brazil*
[e] *CINTESIS – Center for Health Technology and Services Research, Portugal*
[f] *INESCC - Institute for Systems and Computers Engineering at Coimbra, Portugal*
[g] *Faculty of Medicine, Federal University of Rio de Janeiro, Brazil*

Abstract. Data sharing, information exchange, knowledge acquisition and health intelligence are the basis of an efficient and effective evidence-based decision-making tool. A decentralized blockchain architecture is a flexible solution that can be adapted to institutional and managerial culture of organizations and services. Blockchain can play a fundamental role in enabling data sharing within a network and, to achieve that, this work defines the high-level resources necessary to apply this technology to Tuberculosis related issues. Thus, relying in open-source tools and in a collaborative development approach, we present a proposal of a blockchain based network, the TB Network, to underpin an initiative of sharing of Tuberculosis scientific, operational and epidemiologic data between several stakeholders across Brazilian cities.

Keywords. data sharing, interoperability, blockchain, tuberculosis

1. Introduction

Tuberculosis (TB) is a bacterial infectious disease that represents a significant public health problem in the world. It is estimated that in 2017 approximately 10.5 million people got infected with drug-sensitive TB or with drug/multidrug-resistant TB. Brazil, specifically, is considered a country with a high TB burden, with an estimate of 93.000 TB cases [1].

Data sharing is the practice of making data used for scholarly research widely available to other stakeholders [2]. In the absence of any binding requirement, data sharing is at the discretion of the scientists themselves. Data sharing may be restricted to protect institutions and scientists from use of data for political purposes, as well as to

---

[1] Corresponding Author, Filipe Bernardi. Health Intelligence Laboratory, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil; E-mail: filipepaulista12@usp.br.

protect proprietary interests, national security, and subject/patient/victim confidentiality [3].

Data sharing, information exchange, knowledge acquisition and health intelligence are the basis of an efficient and effective evidence-based decision-making tool. This tool needs to be coherent with the institutional and managerial culture of organizations and services. However, current solutions are usually centralized. A valid approach to overcome these risks is to adapt the Blockchain technology where a decentralized architecture allows each member in the network stores an identical copy of the records and contributes to the collective process of validating and certifying digital transactions for the network based on contracts of strict rules and consensus processes [4].

The main goal of this work is to present a proposal of a blockchain based architecture to underpin sharing of TB scientific, operational and epidemiologic data across 7 (seven) Brazilian cities. We intent to build a permissioned network that will allow safe information exchange that could be, ideally, expanded to a global TB network.

2. Methods

Action Research will be used as the scientific methodological basis for this work, due to its investigative and practical approach, considering the project, besides the theoretical development, has also a practical component [5]. All stakeholders will actively participate in the network conception, development and implementation to cover all necessary requirements for a precise data sharing. Documentation will be produced in each development part and a network prototype will be created to validate the proposed model.

The proposed network model of data sharing focus on the construction of structured nodes for a TB project that involves 7 (seven) Brazilian states' capitals, where each city will be able to manage and share information through the blockchain network. The main goal of this network is to provide an approach to effectively and securely share TB information within a data sharing network, called TB-Network.

Several resources will be used to model and deploy the blockchain network for data sharing, which will be based on the open-source Hyperledger platform. Initially, Hyperledger Composer Playground will be used for network modeling, smart contracts definition and testing of transactions. Then, the generated metadata will be exported to be further deployed in a production environment based on the Hyperledger Fabric framework, which will be instantiated in the private academic cloud computing infrastructure from University of São Paulo.

The initial dataset for the blockchain network will came from the Tuberculosis Ecosystem, a computational health infrastructure that consists on a set of integrated systems that aims to do a better management and exchange of information related to TB in State of São Paulo, Brazil. Furthermore, this ecosystem has a functional and semantic interoperability architecture that enables relevant data exchange with authorized systems [6], which will allow data extraction and sharing over the network.

Also, test applications will be developed to interact with the TB Network. In general, it will be web and mobile based applications. They will be used as a proof of concept of supporting tools that can coexist to enhance relevant data dissemination, increasing quality and completeness of specific categories of data.

3. TB Network: a blockchain permissioned network

We have defined general assumptions about healthcare stakeholders (or network nodes) participating in a data sharing network are presented: i*)* Nodes must understand the data structure and semantics; ii*)* Nodes need guarantees of security and auditability to share or receive data; iii*)* The node controls their records and authorizes who may access it and when; iv*)* Each node are able to manage their access levels, users and data privacy. These assumptions exclude any regulation/incentives that the network itself defines.

We propose the use of a permissioned distributed blockchain solution that uses a key pair (private/public key) and a symmetrical consortium key for data encryption. A consortium distributed storage network will be established consisting of research centers and other approved stakeholders throughout the Brazil. Each member organization will undergo an extensive background check. Permissioned access to the blockchain will be granted by a consortium committee, which will assign the keys. Key based mechanisms allow limits to be placed on the data so that administrators, government agencies, and other agencies only have access to the information needed, not the entire records. Additionally, it will be possible to make some pieces of information publicly available, whenever is necessary (for transparency or open data initiatives).

Data will be the digital asset in the blockchain network. Any data generated involving health information can be stored on the blockchain and will be kept "off-chain" in a distributed storage infrastructure. Identifiers and additional hash codes for such data will be stored in the blocks. That way, we will be able to store a large amount of data without overloading the chain but still rely on blockchain security features.

Such data could be, but is not limited to: scientific investigations, medical device data, administrative records, administrative records, socioeconomic information, population studies, recordings generated from IoT devices, medical records, nursing notes, imaging exams, etc. Data will first endure a process of stripping personally identifiable information leading to the generation of two types of data: i*)* Data that is completely stripped of personally identifiable information (this is the data that can be accessed by approved organizations for large scale data analysis and organizational research); ii*)* Data that could not be stripped of personally identifiable information (this information is only available during patient encounters).

Furthermore, there are some projects working in the idea of connecting different Blockchains, where the main goal is to create "The Internet of Blockchains" [7]. It will allow stakeholders to develop additional blockchain networks and end user applications that will be able to interact with the TB-Network (and with its data) and other interconnected infrastructures through a variety of technologies, such as the resources from the Semantic Web. Upon implementation of the network, our next step will be to evaluate its usability and usefulness. To do this, an evaluation protocol, already in discussion with government entities, will be applied. Other ongoing discussions are related to data processing and how to make it available for students and teachers, besides of the establishment of strategies to record the results and to draw conclusions.

4. Final considerations and future directions

We believe that blockchain can play a fundamental role in enabling data sharing within a network. To achieve that, we have defined the high-level structures and protocols necessary to apply this powerful technology to TB issues. The requirements for sharing

data within the healthcare realm are compelling and simply data exchange is not enough. Our proposal aims to demonstrate that effective data sharing networks require consensus on data syntax, meaning, and security of information [8]. It is important to point out that this project is in consonance with the Brazilian government's project that standardizes the opening of public data, whose guiding principles are found in the Brazilian government's data-entry guide [9].

Additionally, besides of the network first implementation, we intent to integrate the presented blockchain infrastructure with two TB projects of high relevance, whose requirements will incorporate new features, such as a set of interconnected models based on Markov and Machine Learning mechanisms, which are incorporated into an evidence-based clinical algorithm (providing recommendations based on WHO guidelines) [10], and a semantic framework that contains an ontology repository and is able to gather health data from heterogeneous data sources [6].

Acknowledgements

References

[1] WHO, Global Tuberculosis Report 2018. 2018.
[2] N.G. Fielding, R.M. Lee and G. Blank, *The SAGE handbook of online research methods*, Sage, 2008.
[3] D.B. Taichman, P. Sahni, A. Pinborg, L. Peiperl, C. Laine, et. al., Data sharing statements for clinical trials: a requirement of the International Committee of Medical Journal Editors, *Jama* **317**(24) (2017), 2491-2492.
[4] T.T. Kuo, and L. Ohno-Machado, Modelchain: Decentralized privacy-preserving healthcare predictive modeling framework on private blockchain networks (2018). arXiv preprint arXiv:1802.01746.
[5] M. Hult and S.A. Lennung, Towards a definition of action research: a note and bibliography, *Journal of management studies* **17**(2) (1980), 241-250.
[6] F.C. Pellison, R.P.C.L. Rijo, V.C. Lima, R.R. de Lima, R. Martinho, R.J.C. Correia, and D. Alves, Development and evaluation of an interoperable system based on the semantic web to enhance the management of patients' tuberculosis data, *Procedia Computer Science* **121** (2017), 791-796.
[7] H. Ugarte, *A more pragmatic Web 3.0: Linked Blockchain Data*, Bonn, Germany, 2017.
[8] V.C. Lima, D. Alves, F.C. Pellison, V.T. Yoshiura, N.Y. Crepaldi and R.P.C.L. Rijo, Establishment of Access Levels for Health Sensitive Data Exchange through Semantic Web, *Procedia computer science* **138** (2018), 191-196.
[9] M.T. Pires, Guia de Dados Abertos - Melhoria do ambiente de negócios por meio da transparência no Estado de São Paulo, 2015.
[10] World Health Organization, WHO treatment guidelines for isoniazid-resistant tuberculosis: supplement to the WHO treatment guidelines for drug-resistant tuberculosis (No. WHO/CDS/TB/2018.7), World Health Organization, 2018.