Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 196 (2022) 434–438

Procedia Computer Science

www.elsevier.com/locate/procedia

CENTERIS - International Conference on ENTERprise Information Systems / ProjMAN - International Conference on Project MANagement / HCist - International Conference on Health and Social Care Information Systems and Technologies 2021

A computational infrastructure for semantic data integration towards a patient-centered database for Tuberculosis care

Vinícius Costa Lima[a,b,*], Filipe Andrade Bernardi[a,b], Michael Domingues[c], Afrânio Lineu Kritski[d], Rui Pedro Chaters Lopes Rijo[e], Domingos Alves[f]

[a] *Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil*
[b] *Bioengineering Postgraduate Program, University of São Paulo, São Carlos, Brazil*
[c] *Faculty of Engineering, University of Porto, Porto, Portugal*
[d] *Medical School, Federal University of Rio de Janeiro, Rio de Janeiro, Brazil*
[e] *School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal*
[f] *Department of Social Medicine, Ribeirão Preto Medical School, University of São Paulo, Ribeirão Preto, Brazil*

Abstract

Tuberculosis is an infectious disease that is among the top 10 causes of death in the world and Brazil ranks in the top 30 high TB burden countries. In this scenario, data integration and sharing are crucial to the construction of efficient and effective evidence-based decision-making tools and to enable data-driven research. Through a socio-technical approach, this work proposes a computational infrastructure composed of a functional and semantic interoperability layer and security mechanisms to integrate national level health information systems towards a patient-centered unified database to provide a broad view of a patient across several isolated databases. The CMIID, a medical information identifier for data harmonization, developed by the University of Porto, was used for the linkage of patients across these health information systems and to perform records anonymization for privacy of personal health information. Through the integration of such systems, it is possible to gather, summarize and visualize TB data in a single system, which can be useful for health professionals and managers. Therefore, this work sought to promote the integration of disparate systems and the availability of data to support decision-making and research, which are fundamental for improving the quality of TB services in Brazil.

* Corresponding author.
E-mail address: viniciuslima@usp.br

1877-0509
10.1016/j.procs.2021.12.033

Keywords: tuberculosis; interoperability; semantic web; health information systems; decision support; data integration

1. Introduction

Tuberculosis (TB) is an infectious disease that is among the top 10 causes of death in the world. Brazil is part of the top 30 high TB burden countries, with an estimate of 96.000 cases in 2019 [1]. It is also a neglected disease [2]. TB health services generally work with few human and technological resources and, often, without the information necessary to deliver adequate health care.

Despite the existence of countless data, some reasons make access by research promotion agencies, academics, managers and health professionals unfeasible, such as non-computerization of processes, heterogeneity and duplicity of data in Health Information Systems (HIS) and the existence of a large amount of isolated data in databases accessible only in a given context. Such factors often cause information quality problems, making it difficult to coordinate and evaluate them in a TB Patient Care Network and impossible to support the decision-making process [3,4].

The global plan to stop Tuberculosis (TB), despite having been approved in 2015 at the World Health Assembly, is currently in a situation where it is estimated, by consensus, that it will be difficult to meet the targets set for 2035 (WHO, 2019). Among the main factors that negatively influence the achievement of goals, it can be highlighted the low quality, availability and integration of data in all phases of TB services and management, as well as their relationship with the organization and management of health services.

There is an urgent need to develop a unified computer information system for the registration, monitoring and evaluation of patients, storing from the most basic information to the data of screening, diagnosis, treatment and hospitalization [5,6]. Also, the integration of data from disparate data sources can contribute to the development of such system, due to existence of relevant data for managerial and operational activities.

To enable this solution, HIS must be able to exchange data with other systems, i.e., they must comply with functional and semantic interoperability standards and protocols. Functional interoperability is the system's ability to communicate in a transparent way. Semantic interoperability is the ability of a given information to be understood by all systems in terms of its formal definition [7].

2. Objectives

This work is being carried out in the municipalities of the 13th Regional Health Department of the State of São Paulo (DRS XIII). The objectives are:

- To develop a computational infrastructure to enable the integration of data available in different systems of health care, social assistance and epidemiology of TB cases, such as the Notification of Injury Information System (SINAN), the Notification and Monitoring System for Cases of Tuberculosis of the State of São Paulo, the Laboratory Environment Manager (GAL), and the Information System for Special Tuberculosis Treatments (SITE-TB);
- To simplify the access of patient data via web by healthcare professionals and managers;
- To establish a patient-centered database to facilitate the query of electronic health records of TB patients in a single system, embracing the whole diagnostic and therapeutic cascade.

3. Methods

A socio-technical approach is being applied in the development, training and to address eventual necessary changes in the solution. In this case, potential users actively collaborate with technical developers, through various strategies, such as regular group meetings, where they are motivated to use and improve the system, integrating their workflows and making suggestions based on their experience [8,9].

The computational infrastructure is composed of three layers, namely: interoperability, security and semantic layer. The interoperability layer holds the webservices that enable the communication and exchange of data between systems. The security layer provides mechanisms for authentication, authorization, confidentiality and integrity of data. Finally, the semantic layer allows the aggregation of meaning in data, including legacy datasets.

The cornerstone of a semantic web-based solution is its underlying ontology [10]. Thus, it is necessary to develop an ontology considering: a) the clinical concepts of TB; b) the several existing systems; c) and TB-related concepts across the diagnostic and therapeutic cascade. The modelling of the semantic structure follows the normalization defined by the World Wide Web Consortium (W3C), so all knowledge generated by the application can be extracted according to the request and reused in other strategic and management information systems [3].

A challenging issue concerns the compatibility of this proposed solution with legacy databases, such as relational databases. To establish a virtual database based on a given ontology and to deliver the capability of running semantic queries on the legacy database, a tool to act as a middleware is needed [11]. The D2R Server (http://d2rq.org/) reduces the impact caused by the paradigm shift from legacy systems to the semantic web. With this approach, there is no need to treat the whole database so that it becomes usable and consumable by web-based semantic applications.

The CMIID, a medical information identifier framework for data harmonization [12], developed by the University of Porto, supports the linkage of patients records across these health information systems. The tool also allows the integration of data based on the Unified Medical Language System (UMLS) semantic groups.

4. Preliminary Results

A data integration strategy and an interoperability solution for the tuberculosis treatment and follow-up in Brazil using semantics technologies and supported by a formal ontology was defined [13]. In this process, Brazilian tuberculosis applications were tagged with entities from the resulting ontology and health professionals can use the data gathered from several data sources to enhance the effectiveness of their actions and decisions.

The CMIID system allows the definition of a preferential patient's identifier for records linkage. However, due to low data completeness, personal identifiers are not always not filled in. Thus, a combination of identifiers (e.g., national ID number, name, mother's name, birth date) is used for the correct identification of patients. In addition to linking records, the tool is also capable of de-identifying records when necessary, depending on the level of access of users.

Additionally, security features were established to protect data shared over the interoperability layer, including a secure channel for encryption and integrity check of transmitted data and the definition of access levels based on ontologies entities [14]. Figure 1 presents the proposed computational infrastructure.

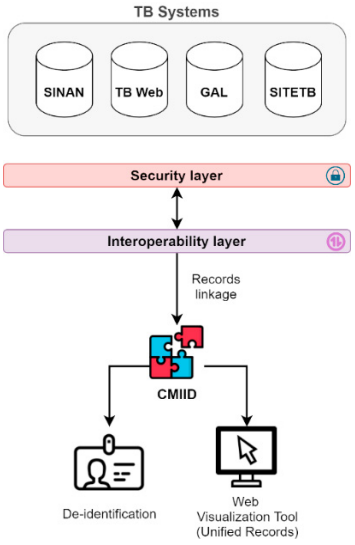

Fig 1. The computational infrastructure for TB data integration and visualization

For testing and validation purposes, epidemiologic data from a regional TB HIS in use in the municipalities of the DRS XIII were integrated with data from the Notification and Monitoring System for Cases of Tuberculosis of the State of São Paulo. In this work, history data were combined and georeferenced for the visualization of TB cases, as shown in Figure 2.

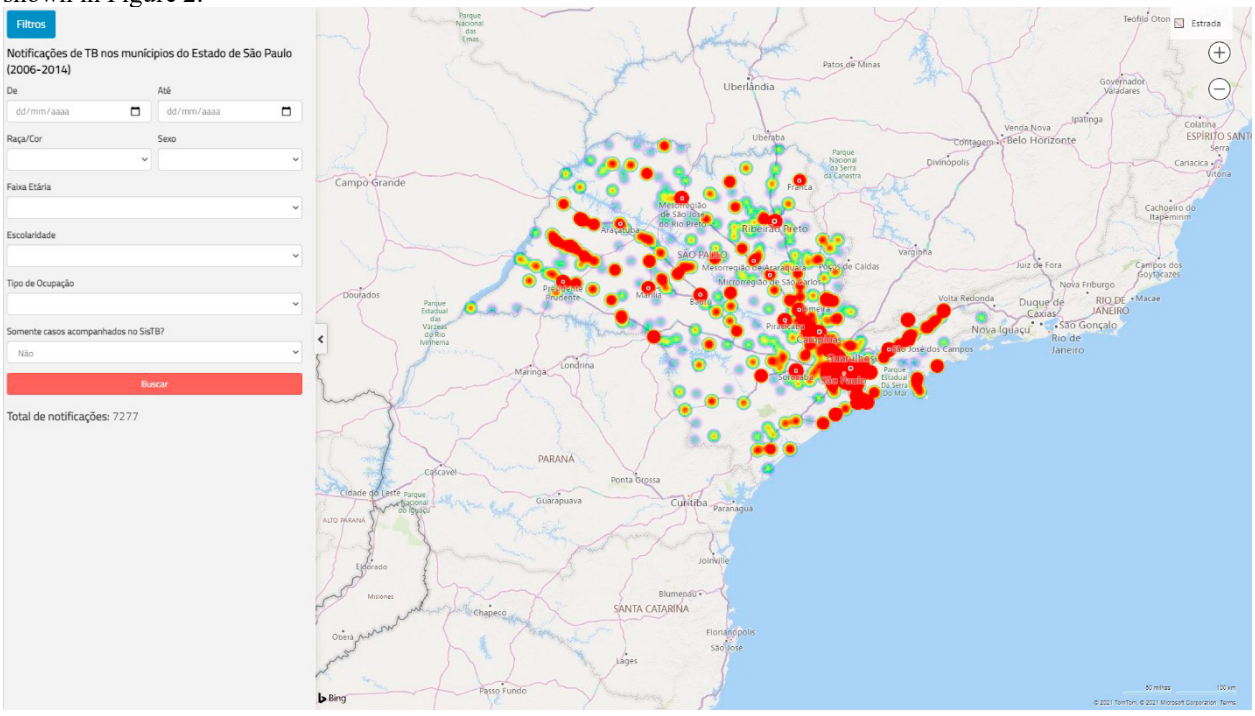

Fig 2. Georeferencing tool – Historical TB cases in the State of São Paulo, Brazil, from 2006 to 2014.

5. Conclusions

This work proposes the development of a computational infrastructure to enable secure functional and semantic interoperability between HIS towards the establishment of a patient-centered database for Tuberculosis care, underpinning decision-making processes of managers and health professionals of Brazilian TB services.

As future work, the development of the Web Visualization Tool will be carried considering the user experience that health professionals and managers demands, which is being defined through the socio-technical approach. It is also intended to gradually implement the solution in real world scenarios. Considering the complexity of each system (candidates for integration) and the data sensitivity, it must be carried out carefully. Also, the effective enrollment of these systems depends on formal agreements with governmental authorities.

Acknowledgements

This study is supported by the São Paulo Research Foundation (FAPESP) - grant number 2020/01975-9, coordinated by author D.A and is under development in the Bioengineering Postgraduate Program (EESC/FMRP/IQSC) of the University of São Paulo.

References

[1] WHO, Global Tuberculosis Report 2020, 2020.
[2] R.G. de Oliveira, Meanings of neglected diseases in the global health agenda: The place of populations and territories, Cienc. e Saude Coletiva. 23 (2018) 2291–2302. https://doi.org/10.1590/1413-81232018237.09042018.
[3] F.C. Pellison, R.P.C. Lopes Rijo, V.C. Lima, R.R. De Lima, R. Martinho, R.J. Cruz Correia, D. Alves, Development and evaluation of an interoperable system based on the semantic web to enhance the management of patients' tuberculosis data, Procedia Comput. Sci. 121 (2017) 791–796. https://doi.org/10.1016/j.procs.2017.11.102.
[4] K. Abhishek, S. M.P., An Ontology based Decision support for Tuberculosis Management and Control in India, Int. J. Eng. Technol. 8 (2016) 2860–2877. https://doi.org/10.21817/ijet/2016/v8i6/160806247.
[5] V.T. Yoshiura, J.M. Azevedo-Marques, M. Rzewuska, A.L.T. Vinci, A.M. Sasso, N.S.B. Miyoshi, A.R.F. Furegato, R.P.C.L. Rijo, C.M. Del-Ben, D. Alves, A web-based information system for a regional public mental healthcare service network in Brazil, Int. J. Ment. Health Syst. 11 (2017) 1–10. https://doi.org/10.1186/s13033-016-0117-z.
[6] World Health Organisation, Providing health intelligence to meet local needs: A practical guide to serving local and urban communities through public health observatories, (2014). www.who.int/about/.
[7] H.Q. Wang, J.S. Li, Y.F. Zhang, M. Suzuki, K. Araki, Creating personalised clinical pathways by semantic interoperability with electronic health records, Artif. Intell. Med. 58 (2013) 81–89. https://doi.org/10.1016/j.artmed.2013.02.005.
[8] A. Luiz, T. Vinci, R. Pedro, C. Lopes, J.M. De, Proposal of an evaluation model for mental health care networks using information technologies for its management, Procedia - Procedia Comput. Sci. 100 (2016) 826–831. https://doi.org/10.1016/j.procs.2016.09.231.
[9] F.C. Pellison, I. Pelogia, M. Damian, R. Pedro, C. Lopes, ScienceDirect ScienceDirect Towards Towards a a health health observatory observatory conceptual conceptual model model based based on on the the semantic semantic web web, Procedia Comput. Sci. 138 (2018) 131–136. https://doi.org/10.1016/j.procs.2018.10.019.
[10] S. Isotani, I.I. Bittencourt, Dados Abertos Conectados : em Busca da Web do Conhecimento, 2015. https://doi.org/10.13140/RG.2.1.4355.6329.
[11] V. Lima, F. Pellison, F. Bernardi, I. Carvalho, R. Rijo, D. Alves, Proposal of an integrated decision support system for Tuberculosis based on Semantic Web, Procedia Comput. Sci. 164 (2019) 552–558. https://doi.org/10.1016/j.procs.2019.12.219.
[12] M.A.P. Domingues, R. Camacho, P.P. Rodrigues, CMIID: A comprehensive medical information identifier for clinical search harmonization in Data Safe Havens, J. Biomed. Inform. 114 (2018) 1–9. https://doi.org/10.1016/j.jbi.2020.103669.
[13] F.C. Pellison, R.P.C.L. Rijo, V.C. Lima, N.Y. Crepaldi, F.A. Bernardi, R.M. Galliez, A.L. Kritski, K. Abhishek, D. Alves, Data Integration in the Brazilian Public Health System for Tuberculosis: Use of the Semantic Web to Establish Interoperability, JMIR Med. Informatics. 8 (2020). https://doi.org/10.2196/17176.
[14] V.C. Lima, D. Alves, F.C. Pellison, V.T. Yoshiura, N.Y. Crepaldi, R.P.C.L. Rijo, Establishment of access levels for health sensitive data exchange through semantic web, Procedia Comput. Sci. 138 (2018) 191–196. https://doi.org/10.1016/j.procs.2018.10.027.
[15] F. Carvalho Pellison, V. Costa Lima, R.P.C.L. Rijo, D. Alves, Integrating Tuberculosis data in State of São Paulo over Semantic Web: a proof of concept, Procedia Comput. Sci. 164 (2019) 686–691. https://doi.org/10.1016/j.procs.2019.12.236.