Available online at www.sciencedirect.com

ScienceDirect

Procedia Computer Science 138 (2018) 131–136

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, CENTERIS/ProjMAN/HCist 2018

Towards a health observatory conceptual model based on the semantic web

Vinicius Tohoru Yoshiura[a,*], Diego Bettiol Yamada[a], Felipe Carvalho Pellison[b], Inácia Bezerra de Lima[c], Ieda Pelogia Martins Damian[d], Rui Pedro Charters Lopes Rijo[e], João Mazzoncini de Azevedo Marques[a], Domingos Alves[a]

[a]*Ribeirao Preto Medical School, University of Sao Paulo, Ribeirao Preto, Brazil*
[b]*Bioengineering Postgraduate Program, University of São Paulo, São Carlos, Brazil*
[c]*School of Nursing of Ribeirão Preto, University of São Paulo, Ribeirão Preto, Brazil*
[d]*Faculty of Philosophy, Sciences and Letters at Ribeirao Preto, University of Sao Paulo, Ribeirao Preto, Brazil*
[e]*School of Technology and Management, Polytechnic Institute of Leiria, Leiria, Portugal*

Abstract

Introduction. Health Observatories have gained global popularity and have been established worldwide since the 1970s. With the advent of the Semantic Web, machines can process, reuse, integrate and understand the meaning (semantics) of the information and knowledge on the World Wide Web to perform complex tasks. Objective. To propose a health observatory conceptual model based on Semantic Web in order to assist in the design, development and implementation processes of a Health Observatory. Methods. The proposed model was based on a conceptual analysis that include Semantic Web technologies narrative review, multi-layer software architecture and an integrative review of Health Observatories. Results. The proposed Health Observatory conceptual model consists of a chain of several related components based on information technology multi-layer architecture, Semantic Web technologies, Health Observatories stakeholders and key concepts. Conclusion. The proposed model can provide opportunities for the development and implementation of new observatories or for the adequacy of existing Health Observatories.

Selection and peer-review under responsibility of the scientific committee of the 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.

Keywords: Health observatory; semantic web; conceptual model

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

1877-0509 © 2018 The Authors. Published by Elsevier Ltd.

Selection and peer-review under responsibility of the scientific committee of the 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.
10.1016/j.procs.2018.10.019

1. Introduction

Health Observatories (HOs) are defined as policy-oriented centres that perform continuous and systematic observations and analyses of health issues relevant to a particular population and a geographic region [1][2][3].

The idea of a HO has gained global popularity since the 1970s because of the rapid changes in the health sector, the growing importance of public health and the need for reliable and up-to-date information for monitoring and evaluating the impact of health programs and policies [4][5].

Typically, a HO has a web portal in order to disseminate and communicate its results in different formats and to generate knowledge and health intelligence in several areas, such as cancer, infectious diseases, mental health and urban health through dynamic and interactive web pages [1][2][3][5].

On the World Wide Web (WWW), the retrieval of information from a large collection of texts is a laborious and time-consuming task for humans and difficult and error-prone for machines [6]. Thus, the concept of Semantic Web (SW) emerged, defined as a network that not only links hypertext documents among themselves, but also recognizes the meaning of these documents and, through a machine processing mechanism, infers new knowledge [6].

Therefore, the objective of this article is to propose a HO conceptual model based on SW in order to assist in the design, development and implementation processes of a HO.

Firstly, we will present a theoretical background on SW and HOs. Secondly, the methods for the elaboration of the conceptual model will be described. Thirdly, we will describe the conceptual model and its mains components as well as discuss the obtained results. Finally, it is presented the conclusion of this paper.

2. Background

In this section, theoretical elements of SW and HO will be presented as the basis of the conceptual model.

2.1 Semantic web

The World Wide Web Consortium (W3C) has been leading initiatives to improve the integration, standardization and exchange of information as well as the semantic understanding of WWW information by machines [7]. One such initiative is the Semantic Web (SW), defined as an extension of the WWW that not only links hypertext documents among themselves, but is also able to recognize the meaning (semantics) of the information contained in these documents and, through inference mechanisms and ontologies, assist in the knowledge management [7].

The SW can be implemented through a combination of technologies, languages and standards recommended by W3C, such as: Resource Description Framework (RDF), a standard language for representing data, metadata and information about resources in the WWW [8]; Resource Description Framework Attribute (RDFa), which provides a set of mark-up attributes to augment the visual information on the WWW with machine-readable hints [9]; Web Ontology Language (OWL), a standard language that can be used to explicitly represent the meaning of terms in vocabularies and the relationships between those terms (ontologies) [10]; and SPARQL Protocol and RDF Query Language (SPARQL), that can be used to perform queries across diverse repositories, in which the data is stored natively as RDF or viewed as RDF via middleware [11]. Furthermore, the Linked Data (LD) approach proposes a set of recommendations and practices that enable applications to expose, query, share and integrate data from different sources through SW technologies as well as to draw inferences using ontologies [7].

2.2 Health Observatory

In general, a Health Observatory (HO) is operationally defined as a policy-oriented centre that performs continuous and systematic observations and analyses of health issues relevant to a particular population and a geographic region in support of evidence-based health policies, decision-making, planning, and actions in health and health systems [1][2][3]. Additionally, a HO is a platform which uses an integrated a set of methods, processes, tools and individuals working together to collect, analyse, report and share data and information and to generate evidence and knowledge to be disseminated to health policymakers, planners and managers of a health system as well as political and health authorities [1][2][12].

The main functions of a HO are related to: provision of a virtual repository for health data and information from different sources; identification of gaps in health information; generation of synthesized, updated and contextualized health reports; provision of services for data and information analysis, interpretation and dissemination; primary commitment to work with partners to support sustainability; and support or recommend evidence-based actions, practices and policies in order to improve health [1][2][3][12][13].

Typically, a HO has functional components and operational resources. The functional components are: inputs, such as data sources and data management; processes, that include production of indicators, knowledge and health intelligence; and outcomes, which are the communication of produced results [1][2][12]. The operational resources, in turn, include: governance; analytical staff; information technology staff; and information and communication technologies [1][2][12].

Recently, some studies have been analysing the use of SW technologies in HOs. In general, they are middlewares for complex data from diverse sources on the WWW. Furthermore, they provide access to visualization and analytical environments capable of collecting, sharing and querying structured and standardized data sets, such as LD, as well as integrating knowledge through ontology management in order to facilitate cross-collaboration analysis for health issues [4][14][15].

3. Methods

Firstly, an integrative review was conducted to investigate the definition, functions, tools and services of a HO. The inclusion criteria were: studies that describe the definition or functions or tools or services of a HO; and studies in English, Spanish or Portuguese. The exclusion criteria were: lack of information (e.g. full text not found); and studies whose focus were not health. The query string 'health observatory' was used and 3456 studies were found on Scopus, 56 on Web of Science, 303 on PubMed, 58 on IEEE digital library, 9 on LILACS and 59 on Worldcat. In addition, 5 WHO and PAHO guidelines and manuals related to HOs were found through handsearching. After the removal of repeated studies and title, abstract and full text analysis, 59 documents, including full articles, editorials, thesis, book chapters and WHO and PAHO manuals were included. Through the findings of this review, it was possible to identify the main definitions of a HO and its key functions, tools and services, which were adopted as the basis of the 'stakeholders' and 'key concepts' components of the conceptual model.

Secondly, a narrative review was conducted in order to explore SW main concepts and technologies. Furthermore, W3C documents for the SW standards were investigated to complement this review [8][9][10][11]. The justification for choosing SW is due to the fact that it is recommended by the W3C and it can, through ontologies and data/information exchange between semantic applications, enhance and assist in knowledge management, which is one of the key concepts of a HO [1][2][12].

Finally, the multi-layer software architecture comprehends different components that are organized in layers in order to provide dedicated functionality and separation of concerns in layers [16]. This architecture along with the findings of the narrative review (SW main concepts and technologies) were adopted as the basis of the 'information technology architecture' component in the conceptual model.

4. Results and discussion

The proposed HO conceptual model consists of a chain of several related components based on information technology multi-layer architecture, SW technologies, stakeholders as well as key concepts (Fig. 1).

The 'stakeholder' component represents the stakeholders of a HO and their roles. It is composed by: information technology (IT) staff, responsible for IT activities; analytical staff, responsible for data-related and analytical tasks; end users, such as health professionals, policy makers, planners, managers, political and health authorities, and researchers; governance staff, responsible for coordinating HO internal staffs, guaranteeing sustainability and establishing partnerships; and SW applications, which communicate with the 'IT architecture' component through LD. In addition, the governance staff needs to work periodically and frequently with end users in order to validate whether the HO functionalities and operations are coherent and responsive to their needs [1][2].

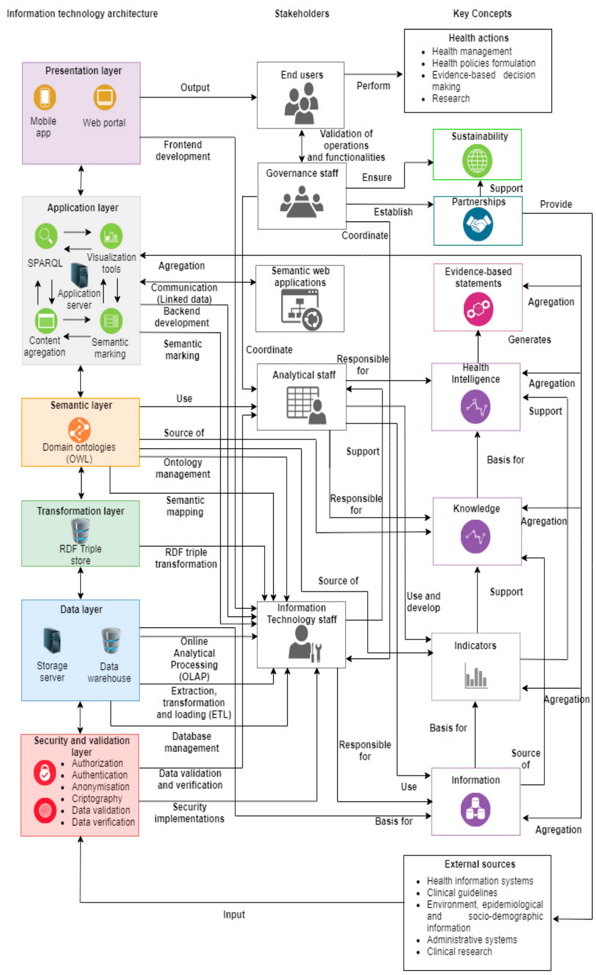

Fig. 1. Health observatory conceptual model based on the semantic web.

The 'information technology architecture' component is based on the multi-layer software architecture and SW technologies. The main idea of this component is to represent, in different layers, the information technology artefacts and processes needed for data/information security, management, integration, analysis, dissemination and how the SW technologies can be linked into this structure. It has the following layers:

- Security and validation layer: the IT staff is responsible for ensuring data security from different external sources through authorization, authentication, anonymisation and cryptography mechanisms, while the analytical staff is responsible for performing data validation and verification. Concerns about data security is an essential and sensitive part of a HO in order to preserve the confidentiality and privacy of patients, health professionals and institutions [17]. Furthermore, when the set of data incorporated in a HO is treated and validated, it can be appropriately used in evidence-based decision making and in the development of health policies and guidelines [1];
- Data layer: after the first layer, an Extraction, Transformation and Loading (ETL) process is applied to integrate the datasets from external sources and, subsequently, are uploaded into a data warehouse. Additionally, an Online Analytical Processing (OLAP) is used to generate information, process sets of indicators and provide support to data analysis. The importance of adopting a data warehouse is related to its capability of managing large amounts of data in an easily, accessible, consistent and adaptive way [1][2][18]. Within the OLAP engine, it is possible to create and manage aggregations through precalculations, indexing strategies and other optimizations [18];

- Transformation layer: transforms the stored data from data layer into RDF triples. The set of RDF triples can be stored into RDF triple store, a database which is structured and optimized for storage and retrieval of triples [19]. The process of RDF transformation involves extracting the OLAP results from data layer, transform them into a triple format and notation and load the resulting triples into the RDF triple store;
- Semantic layer: contains a set of ontologies represented in OWL. The resulting RDF triples from the transformation layer can be mapped into a set of ontologies using OWL. Additionally, developed ontologies can be added into the HO set of ontologies, requiring an ontology management [7][9][15]. Hence, in order to assist in the integration of knowledge derived from a variety of ontologies and to provide a common ontological architecture, a standard top-level ontology, such as the Basic Formal Ontology (BFO), is necessary [15][20];
- Application layer: includes the business logic of the HO. The set of ontologies from Semantic layer, whose content is stored in RDF, can be queried through SPARQL, enabling data, information, indicators and knowledge aggregation across diverse SW applications, RDF repositories and related ontologies in OWL [19]. The content semantic marking is possible through the RDFa, allowing other SW applications to process and understand the content that is made available [9]. In addition, visualization tools aim to prepare the results of the SPARQL queries and the knowledge produced by the analytical staff for the presentation layer;
- Presentation layer: it is an accessibility layer to allow end users to access the results from application layer in a structured manner through a web portal or a mobile app. All user interfaces styles and dynamics need to be implemented by IT staff (frontend development). This layer depends on the proper functioning of all others layers in order to support the generation and visualization of HO main outputs, based on SW technologies, such as: summaries of the health situation and trends; analytical reports; indicators, dynamic tables, charts, interactive maps and geographic variations; and evidence-based, technical and methodological recommendations, comments and guidelines [1][2][3][5].

The ‘key concepts’ component represents the main concepts and functionalities for the HO operation. Firstly, external sources provide the input for the ‘IT architecture’, in which health information systems, administrative systems, environment, epidemiological and socio-demographic information offer the healthcare primary datasets, through direct database import or Application Programming Interfaces (APIs) [21]. Moreover, clinical researches, such as randomized controlled trials and systematic reviews, are required for generation of knowledge, health intelligence and evidence-based statements [1]. Secondly, after the data layer operations, the information is produced from ETL and OLAP processes. Subsequently, the knowledge can be generated though analysis by the analytical staff based on these produced information along with indicators and SW ontologies. The next step is to translate knowledge into health intelligence, which is produced when the knowledge is analysed by specialized and experienced members of the analytical staff from different perspectives and is compared with clinical research and guidelines. Then, the health intelligence can be structured to generate evidence-based statements [1]. Thirdly, end users use the output from the web portal or the mobile app to perform health actions, such as health management, health policies formulation, evidence-based decision making and research [1][2][3][12]. Finally, the governance staff needs to maintain and expand a network of partnerships, which is necessary to guarantee access to external sources, learn from other established HOs and ensure the HOs sustainability [22].

The proposed HO conceptual model aims to assist in design, development and implementation processes of an HO. Therefore, it is expected that the model can be replicable and valuable, as well as flexible enough to be used in different health domains to encourage creativity and novel use [17]. The use of SW technologies can enhance the HO in terms of data standardization, knowledge management and interoperability with other SW applications. However, in order to be sustainable and successful, a HO need to: have an integrated, specialized and committed staff (IT, analytical and governance); have the involvement and feedback from local community and key stakeholders; be responsive to national, regional, and local information priorities; take into consideration that the development process of the HO is a labour-intensive process that takes time and planning; have a careful strategic project monitoring and periodic evaluation; and consider that the outputs of the HO depend on the availability and the development level of external sources [1][2][5][12][22].

5. Conclusion

The proposed HO can provide opportunities for the development and implementation of new observatories or for the adequacy of existing HOs. Consequently, these opportunities may considerably contribute to the health policies formulation, health management, evidence-based decision making and health research, thus, improving, preserving and promoting health and quality of life of the population.

The next steps involve a practical application of this model in a specific health domain, which requires exploring and analysing the specific context in which the HO will be developed, through the identification of end users problems and needs, relevant external sources to be the HO inputs, as well as partnerships, financial and human resources in order to ensure the HO sustainability.

References

[1] Rodrigues RJ, Gattini, CH (2016) "National Health Information Systems and Health Observatories**",** in: Marin, HF et al. "*Global Health Informatics How Information Technology Can Change Our Lives in a Globalized World*", Amsterdam, Netherlands, Elsevier.

[2] Gattini CH (2009) "*Implementing National Health Observatories*", Santiago, Chile, Pan American Health Organization.

[3] World Health Organization (2014) "*Providing health intelligence to meet local needs: a practical guide to serving local and urban communities through health observatories*", Kobe, Japan, World Health Organization.

[4] Zaveri A et al. (2013) "Publishing and interlinking the Global Health Observatory dataset" *Semantic web journal* **4** (3): 315-322.

[5] Aspinall PJ, Jacobson B, Castillo-Salgado C. (2016) "Establishing and sustaining health observatories serving urbanized populations around the world: scoping studying and survey" *The European Journal of Public Health*, **26** (4): 681-686.

[6] Berners-Lee T, Hendler J, Lassila O (2001) "The semantic web" *Scientific American* **248** (5): 34-43.

[7] Isotani S, Bittencourt I (2015) "*Dados abertos conectados*", São Paulo, Novatec.

[8] Manola F, Miller E (2004) "RDF Primer" [accessed in april 11, 2018]. Available on: https://www.w3.org/TR/2004/REC-rdf-primer-20040210/]

[9] Herman I et al. (2015) "RDFa 1.1 Primer - Third Edition" [accessed in april 27, 2018]. Available on: https://www.w3.org/TR/rdfa-primer/

[10] McGuiness DL, van Hermelen F (2004) "OWL Web Ontology Language Overview" [accessed in april 11, 2018]. Available on: https://www.w3.org/TR/2004/REC-owl-features-20040210/#s1.1

[11] Prud'hommeaux E, Seaborne A (2008) "SPARQL Query Language for RDF" [accessed in april 11, 2018]. Available on: https://www.w3.org/TR/rdf-sparql-query/

[12] World Health Organization (2016) "*Guide for the establishment of health observatories*", Brazzaville, Republic of the Congo, World Health Organization.

[13] Hemmings J, Wilkinson J (2003) "What is a public health observatory?" *Public health policy and practice* **57** (5): 324-326.

[14] Tiropanis T et al. (2014) "The web observatory: A middle layer for broad data" *Big Data* **2** (3): 129-133.

[15] Shaban-Nejad A et al. (2017) "PopHR: a knowledge-based platform to support integration, analysis, and visualization of population health data" *Annals of the New York Academy of Sciences* **1387** (1): 44-53.

[16] Schuldt H (2009) "Multi-tier architecture", in: Liu L, Özsu MT "*Encyclopedia of Database Systems*" New York, United States of America, Springer.

[17] Gloria MJKT et al. (2014) "Legal and Ethical Considerations: Step 1b in Building a Health Web Observatory" Proceedings of the 23rd International Conference on World Wide Web, Seoul, Korea.

[18] Kimball R, Ross M (2002) "Data Warehousing, Business Intelligence, and Dimensional Modeling Primer" in: Kimball R, Ross M "*The data warehouse toolkit: the complete guide to dimensional modelling*", Indianapolis, United States of America, John Wiley & Sons.

[19] Allemang D, Hendler J (2011) "*Semantic web for the working ontologist: effective modeling in RDFS and OWL*." Amsterdam, Netherlands, Morgan Kaufmann.

[20] Arp R, Smith B, Spear AD (2015) "*Building Ontologies with Basic Formal Ontology*" London, England, MIT Press

[21] Yoshiura VT et al (2017) "A web-based information system for a regional public mental healthcare service network in Brazil" *Int J Ment Health Syst* **11**: 1

[22] Caiaffa WT et al (2014) "Development of a Conceptual Framework of Urban Health Observatories Toward Integrating Research and Evidence into Urban Health Policy for Health and Health Equity" *Journal of Urban Health: Bulletin of the New York Academy of Medicine* **91** (1): 1-16.