A Data Integration Model for a Decision Making System in Health

Odara Sena
Mestrado de Computação Aplicada
Universidade Estadual do Ceará
Fortaleza, Brazil
odara.sen@gmail.com

Mauro Oliveira
Laboratório de Redes de Aracati
Instituto Federal do Ceara, IFCE (Unidade Aracati)
Aracati, Brazil
amauroboliveira@gmail.com

***Abstract*—LARIISA is an intelligent system to support decision-making in health governance. It uses context-aware concept, ontology and tracking technologies to support the achievement of inference by the system. This article introduces a data integration platform for the Lariisa, where it have as main objective to allow the integration of various databases relevant to health, with various governance issues involved, enabling a higher quality on the inferences made by Lariisa. Such integration is achieved through the Linked Data Mashups that will be created according to each demand of inference made to the Lariisa.**

Keywords— Linked Data; ontology; e-health; Semantic Web.

I. INTRODUCTION

Despite the significant progress that Brazil has done in the area of health over the years, there remain challenges related to inefficiencies and poor quality of the services provided [1]. Many of these challenges are linked to failures of governance. Governance, according to the World Bank, is the manner in which power is exercised in the administration of social and economic resources of a country aiming at the development, and the capacity of Governments to plan, formulate and carry out policies and scheduling functions.

In this sense, information technology shows itself as a mainstay for obtaining this governance. In the area of health, for example, information technology acts as a facilitator of the processes of health providers, provides tools that help healthcare professionals to make the most accurate diagnosis and can also assist in the management of services provided.

The Lariisa project [2] has been proposed to perform health information's inference to assist decision making health Manager. The platform is based on context-aware health applications, i.e., that take into account information that may be used to characterize the situation of an entity [3]. Means that the inferences made by the Lariisa can be obtained from information generated in a short period of time and not only from a long history of information.

Besides the advantage of being context-aware, the Lariisa proposes to accomplish these inferences from health information that represent 5 areas of intelligence, called the Governance Health System, which according to [4], they are needed for a Healthcare Manager make the best decisions. They are: (1) intelligence of knowledge management, (2) normative intelligence, (3) clinical-epidemiological intelligence, (4) administrative intelligence, and (5) shared intelligence.

With that, the work designed in this article constructs a model of how the data should be submitted to Lariisa, whether these data come from applications built exclusively to provide information to the platform, whether data coming from existing health systems to provide data for the Lariisa, whether open health data already available. The proposed model also specifies how the Lariisa will perform inferences from these data obtained.

The Semantic Web is distinguished in the construction of this work. She provides technologies to effectively publish, retrieve and describe data distributed on the Web. The large-scale data integration is probably one of the best use cases for Semantic Web technologies [5], [6]. The Semantic Web considers the concept of migrating the "Web of documents" to a "Web of data", so instead of the cluster of pages and links between them that we have today, we will have a mass of data, interconnected, available in the Internet.

This data integration in the Semantic Web is achieved through the use of Linked Data. Linked Data is a set of principles and best practices that are based on Semantic Web technologies such as RDF and SPARQL, and reduces the complexity of data integration due to the links correctly established and described between sources [7], [8].

Given the above, the model built integrates health data, using the technologies and recommendations of the Semantic Web and Linked Data. Such linked information can be consumed, manipulated and used by Lariisa.

The main contribution of this work is the possibility to implement the main concept of Lariisa, that is the realization of information inference based on 5 management intelligence domains already mentioned, since previous work about the Lariisa were focused on applications that provided information only on an area of specific intelligence, the clinical-epidemiological intelligence.

II. THE LARIISA PLATFORM

Applications developed for the Lariisa must be able to detect the change in context of the environment in which they

find themselves and provide information so that it is possible to perform inferences to aid in the decision-making process of Health Manager.

The health context to which the Lariisa is centered [9] characterizes situations of entities such as family members, a health officer, the health Manager, among others, i.e. entities considered relevant for the interactions between a user and a health system.

For this purpose, the LARIISA defines two ontologies for information modeling of local and global health context. Local health context describes the situation of any entity interacting with the system of governance, such as end users (patients), health managers, health agents, etc. This information is used for the definition of rules of local health and decision to build global health context that describes high-level information, derived from local health context, and is used for decision-making in health governance. For example, the global health context describes the number of cases of dengue fever confirmed in a region (e.g., neighborhood, city, community), over a given period of time (e.g., one day a week). Therefore, this information may be seen as global indicators used to improve governance decisions.

The Lariisa stands out against other intelligent systems in the area of health management, because it relies on the Governance Health System by [4], as already mentioned. In Figure 1, we have the conceptual model of the Lariisa, performing inference of information from 5 domains.

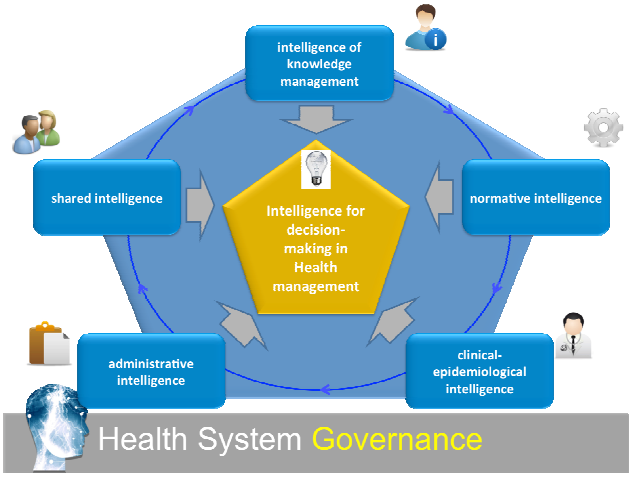

Fig. 1. Domains of intelligence in Health Governance

In practice, the Lariisa may have applications developed with the purpose of providing information concerning a domain specific intelligence, as did [10] who develop a prototype that had set the aggravations of dengue fever in the State of Ceará, and [11] who proposed an environment of integration of heterogeneous context providers, such as Digital TV decoders, mobile devices sensing equipment, etc. Both developed their work considering the clinical and epidemiological domain.

More recently, in the work of [12], the LARIISA added in its local context ontology concepts of Geographically Encoded Objects for RSS (Really Simple Syndication) feeds called GeoRSS, a simple markup with location information to coordinate description and geo spatial relations, which is used to represent temporal content, enriching the patient health data collected by a mobile device, for example.

III. Related Work

Some earlier work dealt with data integration with Linked Data. In [13], DIGO (Delivery Information of GOvernment) is proposed. It allows primary governmental data access by machines in the form of open data enabling interested citizens have access to this data and can combine them and produce new information and Mashup[1] applications.

The work of [14] also contributes to the development of this work. Reference [14] described an efficient mechanism for integration of data on Linked Data from an architecture of Linked Data Mashups based on the use of Linked Data Mashup Services. Developed a module for efficient execution of federated queries over Linked Data.

The unpublished work of [15] was extremely significant, because in his work he presented a Linked Data framework able to make transparent the process of inclusion of new data sources. As a case study of his work he joined various data sources to confront municipal purchases made by the Government.

Another very relevant work was the [16] one. In this work the authors demonstrate the advantages in the use of Linked Data for pharmacological data integration. With the connection of some sets of data available in the Linked Open Data (LOD) cloud[2], the case study presented allows the business manager navigate through data relating companies, clinical trials, drugs, diseases and genetic variations.

IV. Metodology

The health data integration proposal in this paper assumes that the datasets to be used are numerous and heterogeneous, and it are be not considered only data sets already available on the LOD cloud, but it is possible to integrate data that are not yet in the Linked Data model. Thus, the LARIISA shall provide the most reliable information to the Health.

For the model construction, we chose make use of the mediation framework, based on ontologies, proposed by [17]. He uses the concepts of application ontology that represents the common vocabulary among the various sources of data, and exported ontologies, which model each data sources used.

The case study implemented takes into account the data integration from two areas of intelligence in management: clinical-epidemiological intelligence and shared intelligence. We made the integration of the data relating to dengue cases in the city of Fortaleza, representing the clinical-epidemiological intelligence, and with the data of the Observatory of Dengue[3] representing the shared intelligence. The data from the Observatory will show in real time the feeling of the Brazilian

1 Mashups are new development approaches where aggregate multiple services for the creation of a new service to serve a new purpose.

2 Set of Linked Data published by community Linking Open Data or by individuals or organizations in isolated initiatives.

3 www.observatorio.inweb.org.br/dengue/

population about dengue fever via Twitter, and can be considered a good source of epidemiological information [18], [19]. Only with the integration of information belonging to two intelligence domains we realize the wealth of inferred decisions compared to inferences made with information from only one domain.

The goal in integrate these two data sources is to provide to the health manager dengue aggravations information of Fortaleza city arising both from official sources, getting the data from the Monitoring System of Daily Aggravations (SIMDA)[4], as unofficial sources, which is the case of Dengue Observatory. Thus, it is possible to obtain information more expressive, seen the possibility of outbreak information in real time, before even notifying performed in hospitals.

Next, we define a process for integration. Thus, in general terms, we define the proposed integration model for carrying out the following steps:

STEP 1: modeling application ontology after defined the health set information that will be delivered by Lariisa. In the case study of this paper the health information delivered by the LARIISA is related to the dengue aggravations, then our Application Ontology should have a vocabulary related to this.

STEP 2: selection of data sources to be integrated and modeling of Ontologies Exported, based on information that you want to deliver. The sources may already be in the Linked Date format or be in a relational database. If the data is in the relational form, which proceeds on the presented case study, we must perform the data triplication according to the ontology exported defined for the data source.

STEP 3: Heuristics to identify sameAs links between ontologies. In this step, performed a priori manually, should be identified similar vocabularies between the different ontologies exported. This is an important step because it represents exactly the links between sources.

STEP 4: Populate the ontology exported, wiping data with conflicting values.

V. Results

With the development of this case study, where it joined two distinct data sources, an official source of dengue notification in the Fortaleza city, and another data source, considered unofficial, from shared information on Twitter, the LARIISA offers to the Health Manager more complete information about the situation of dengue fever in the city.

As the Observatorio da Dengue project has 85% reliability, this LARIISA application enables the Manager to identify in real time the number of outbreaks.

The interface of access to information is still available in the form of a SPARQL endpoint, but can easily be developed a more intuitive interface, with the approach of maps, for example.

4 Monitoring system of aggravations of dengue developed by technicians of the cell for the epidemiological surveillance of Fortaleza Municipal Health Secretariat - http://tc1.sms.fortaleza.ce.gov.br/simda

From the construction of this case study, we use it as a model for the development of new applications LARIISA built from the integration of different data sources, whether these sources are already triplificadas or in the relational form. With the use of this template, we can standardize the creation of new applications LARIISA, as shown in Figure 2. Figure 2 explains the LARIISA as being the sum of his core more applications that will be developed for the platform.

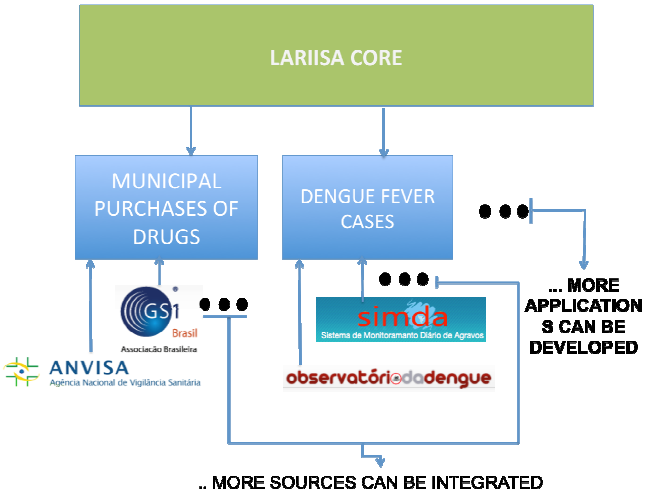

Fig. 2. Development of applications with integrated data for LARIISA

VI. Conclusion

This article presented a data integration model for an intelligent system of health, the LARIISA platform. The model was built from the development of a case study. The case study dealt with an application LARIISA whose goal is to present the situation of the aggravations of dengue in the Fortaleza city, relying on data integration of SIMDA and Dengue Observatory, getting a most accurate information about the outbreaks of dengue fever.

With the development of LARIISA applications that integrate data sources related to the domains of intelligence needed for a health Manager take their decisions, as postulated by [4], the LARIISA can have access to all this data, performing thus the inferences in health as expected.

References

[1] [1] BANCO MUNDIAL, “Brasil Governança no Sistema Único de Saúde (SUS) do Brasil : Melhorando a Qualidade do Gasto Público e Gestão de Recursos,” 2007.

[2] [2] M. Oliveira, C. Hairon, O. Andrade, R. Moura, C. Sicotte, J.-L. Denis, S. Fernandes, J. Gensel, J. Bringel, and H. Martin, “A context-aware framework for health care governance decision-making systems: A model based on the Brazilian Digital TV,” *2010 IEEE Int. Symp. “A World Wireless, Mob. Multimed. Networks,”* pp. 1–6, Jun. 2010.

[3] [3] A. K. Dey, “Providing Architectural Support for Building Context-Aware Applications Providing Architectural Support for Building Context-Aware Applications,” Georgia Institute of Technology, 2000.

[4] [4] L. O. M. de Andrade, “Inteligência de Governança para apoio à Tomada de Decisão,” *Cien. Saude Colet.*, vol. 17, no. 4, pp. 829–832, Apr. 2012.

[5] [5] R. P. MAGALHÃES, “UM AMBIENTE PARA PROCESSAMENTO DE CONSULTAS FEDERADAS EM LINKED DATA MASHUPS,” UNIVERSIDADE FEDERAL DO CEARÁ, 2012.

[6] [6] W3C BRASIL, “W3C Brasil - World Wide Web Consortium Escritório Brasil,” 2011. [Online]. Available: http://www.w3c.br/Padroes/WebSemantica. [Accessed: 15-May-2013].

[7] [7] F. Bauer and M. Kaltenböck, *Linked Open Data : The Essentials*. 2012.

[8] [8] T. Berners-Lee, “Linked Data - Design Issues,” 2006. [Online]. Available: http://www.w3.org/DesignIssues/LinkedData.html. [Accessed: 28-Nov-2012].

[9] [9] J. H. Jahnke, Y. Bychkov, D. Dahlem, and L. Kawasme, “Implicit , Context-Aware Computing for Health Care,” in *Proceedings of the 1st Int’l Workshop on Modeling and Retrieval of Context (MRC 2004). Schulz, S. and Roth-Berghofer, Th.(Eds.), CEUR Workshop Proceedings.*, 2004.

[10] [10] F. Antunes, “SISA – Uma aplicação sensível ao contexto para agravos de Dengue : uma prova de conceito do projeto Lariisa,” 2011.

[11] [11] J. B. B. Frota, “Proposta de solução de integração de provedores de contexto ao sistema lariisa,” 2011.

[12] [12] L. M. Gardini, R. Braga, J. Bringel, C. Oliveira, R. Andrade, H. Martin, L. O. M. Andrade, and M. Oliveira, “Clariisa, a context-aware framework based on geolocation for a health care governance system,” *2013 IEEE 15th Int. Conf. e-Health Networking, Appl. Serv. (Healthcom 2013)*, no. Healthcom, pp. 334–339, Oct. 2013.

[13] [13] A. L. Machado and J. M. P. de Oliveira, “DIGO: An Open Data Architecture for e-Government,” *2011 IEEE 15th Int. Enterp. Distrib. Object Comput. Conf. Work.*, pp. 448–456, Aug. 2011.

[14] [14] R. P. MAGALHÃES, M. S. Maia, V. M. P. Vidal, J. A. F. de Macedo, J. M. Monteiro, and F. A. M. Porto, “Desenvolvimento de Linked Data Mashups com o uso de LIDMS,” in *Simpósio Brasileiro de Bancos de Dados - SBBD 2012*, 2012, pp. 217–224.

[15] [15] R. G. Mariano, M. Maia, and D. Miranda, “CONSTRUINDO UM MASHUP PARA INTEGRAR DADOS DE COMPRA MUNICIPAIS PUBLICADOS NA WEB VIA LINKED DATA,” in *V Encontro de Pesquisa e Pós-Graduação da UFC*, 2012.

[16] [16] A. Jentzsch and O. Hassanzadeh, “Enabling tailored therapeutics with linked data,” in *2nd Workshop on Linked Data on the Web (LDOW2009)*, 2009, pp. 1–6.

[17] [17] J. C. Pinheiro, “Processamento de Consulta em um Framework baseado em Mediador para Integração de Dados no Padrão de Linked Data,” Universidade Federal do Ceará, 2011.

[18] [18] D. M. da S. Barros, P. S. G. de Morais, J. C. de Paiva, J. R. F. de Lima, and J. L. R. da Silva, “OBSERVATÓRIO NACIONAL DA DENGUE - SISTEMA PARA MONITORAMENTO DE CASOS DE DENGUE,” *Rev. Bras. Inovação Tecnológica em Saúde ISSN 2236-1103*, vol. 3, no. 4, Feb. 2014.

[19] [19] J. Gomide, A. Veloso, W. Meira, V. Almeida, F. Benevenuto, F. Ferraz, and M. Teixeira, “Dengue surveillance based on a computational model of spatio-temporal locality of Twitter,” in *Proceedings of the 3rd International Web Science Conference on - WebSci ’11*, 2011, pp. 1–8.