

A Change Management Dashboard for the SIEMA Malaria Surveillance Infrastructure

Jon Haël Brenas¹, Mohammad Sadnan Al-Manir², Christopher J. O. Baker², and
Arash Shaban-Nejad¹

¹ The University of Tennessee Health Science Center- Oak Ridge National Laboratory
Center for Biomedical Informatics, Department of Pediatrics, Memphis, Tennessee, USA
{jhael, ashabann}@uthsc.edu

² Department of Computer Science, University of New Brunswick, Saint John, Canada
{sadnan.almanir, bakerc}@unb.ca

Abstract. Malaria is an infectious disease that remains a major cause of death in low-income developing countries. The World Health Organization (WHO) has set a target for its eradication by 2030. Among the issues that will have to be solved to achieve this goal is interoperability between the various malaria data sources. This can be achieved through the adoption of semantic web service infrastructure to provide access to the data while abstracting its structure. Given that data sources, semantic metadata descriptions and ontologies evolve over time, it remains a challenge to propagate changes, ensuring services continue to be discoverable, while at the same time keep the services operational. We propose a dashboard to detect, identify, and classify changes based on their likely functional impact on data access, and propose steps to maintain infrastructure, either rebuilding or retiring services from a registry.

1 Introduction

Malaria is an infectious disease caused by a parasitic microorganism and considered as one of the major causes of death in low-income developing countries (LIDCs) [1]. In 2015 [2], there were 212 million new cases of malaria, and more than 4 million malaria deaths, worldwide. African countries suffer most, bearing almost 90% of the burden in terms of global cases and deaths. In order to control and ultimately eradicate malaria, access to an efficient real-time surveillance based on consistent integrated data sources is of utmost importance. These data sources contain different types of data ranging from drug efficacy to patient profiles, vector species, climate, and location specific infection data. These heterogeneous data sources are distributed globally. The Semantics, Interoperability, and Evolution for Malaria Analytics (SIEMA) project aims to use ontologies (e.g. IDOMAL [3]) to integrate knowledge from domains of interest and combine malaria specific data or information from multiple heterogeneous sources as well as support interoperability between stakeholder organizations within and across different enterprises and geographies.

Given the complexities of working with dynamic evolving ontologies [4] and databases, a semantic middleware solution in support of non-technical surveillance practitioners seeking to discover and query target information is appropriate. For this purpose we have adopted SADI (Semantic Automated Discovery and Integration) [5], an existing standards-compliant Semantic Web service design pattern that simplifies the publication of services. Specifically, Web service input, output, and functionality are described using the Web Ontology Language (OWL) and access to data is provided using the HTTP-based recommendations (GET, POST). SADI services in a registry can be discovered and queried using dedicated clients like SHARE [6].

In addition to leveraging semantic middleware in support of interoperability between various sources of data, a viable solution must be robust even when data sources and access are modified. When the structure of a database, a domain ontology or a service ontology changes, service descriptions may become inconsistent with the data services they represent, leading to a state where the service can no longer serve the expected data. To report on service uptime and integrity we propose a dashboard to detect and report on changes in service infrastructure, identify and classify changes and propose further steps to maintain the integrity of SIEMA by either rebuilding the services using ValetSADI [7] or retiring them. The dashboard is designed for use by technical staff responsible for malaria surveillance tasks. In the following sections, we discuss the design of the dashboard, tools for detecting changes, the types of changes reported and the use of individual widgets for each case.

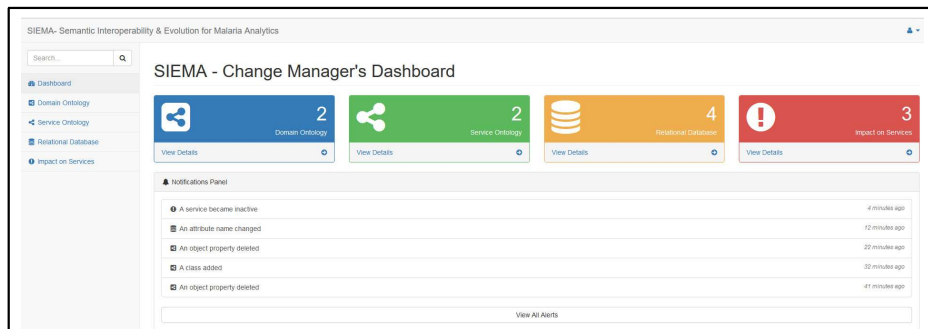


Fig 1. The SIEMA Change Management Dashboard.

2 Dashboard

Our proposed dashboard is composed of four widgets, three of which represent the potential sources of change while the last one contains the list of services. We will discuss these widgets in the following subsections. The current version can be accessed at <http://cbakerlab.unbsj.ca:8080/siemadashboard/pages/index.html>.

In our design, changes are classified into effectual and ineffectual changes. Ineffectual changes do not modify the knowledge (e.g. adding that *m* is an *insect*

when it is known that *m* is a *mosquito* and that all *mosquitos* are *insects*). On the other hand, an effectual change modifies the knowledge.

2.1 Domain ontology

The Domain Ontology page is represented in two panels that show the current and previous versions of the ontology side by side. Tools like ecco [8] can show the differences between two versions of an ontology classifying them into effectual and ineffectual changes.

2.2 Service ontology

The Service Ontology page shows the current and previous versions of the service ontologies that describe the services that are available. They can be modified reactively when changes occur in domain ontologies or in their dependent artifacts (e.g. databases) and registries of service ontologies can be updated accordingly. Since the difference between a domain ontology and a service ontology is in the semantics and not in the syntax, ecco can be used to identify effectual and ineffectual changes in service ontologies as well.

2.3 Databases

The Databases page displays the current and previous version of a database. The databases evolve more frequently because data is constantly added or removed from them. Databases also change when their schemas are modified. The Databases page makes use of tools like Liquibase [9] that express changes in the database in an XML-like format.

2.4 Impact on Services

The Impact on Services page displays the list of services and reports their status. Active services, services that are ready to use, are displayed in green, while inactive services, services that do not work, are displayed in red. To resolve inconsistencies between domain and service ontologies we propose that rules be defined that update obsolete services so that they can still be used. For example if a service is described as `getGeoLocationbyMosquitoSpeciesName`, and a change occurs in the domain ontology, where the concept *mosquito* is replaced with the concept *culicidae*. A rule would then replace all occurrences of *mosquito* in the service definitions have to be replaced with *culicidae* and the service would become `getGeoLocationby-CulicidaeSpeciesName`. Further evaluation of precise inputs and outputs would also need to be made. Additionally services may be regenerated de novo using Valet SADI in cases where source data schemas have been updated.

3 Discussion

Being able to work with heterogeneous and distributed data sources and ontologies is crucial in many public health applications that demand timely response. We employed Semantic Web services to improve semantic interoperability between malaria data sources. One of the challenges is how to maintain their integrity when their descriptions, the data they access or their related ontologies evolve over time. We propose a dashboard that enables us to detect, identify and handle such changes. However, there remains a number of theoretical challenges that need to be addressed. Among the most important is the definition of the rule templates that would make possible the automatic update of Semantic Web services according to changes in the ontology or the database schemas. In order to be able to define those templates, a profound theoretical and empirical study of possible changes and their impact on services is required. Also, the dashboard is currently acts as a reactive reporting tool. It means that it reacts to changes after they are propagated in the system but it does not permit its users to preview the consequences of non-executed changes. In the future we plan to make the dashboard more proactive by enabling service providers, database and ontology managers to propose changes and check whether the consistency of the whole system is preserved.

Acknowledgements

Research supported by the Bill and Melinda Gates Foundation.

References

1. WHO: The top 10 causes of death, 2015.
2. World Health Organization. World malaria report 2016. Technical report, 2016.
3. Wilkinson, M. D., Vanderwalk, B., McCarthy, L.: The Semantic Automated Discovery and Integration (SADI) web service design-pattern, API and reference implementation. *Journal of Biomedical Semantics* 2(1) (2011) 8
4. Topalis, P., Mitraka, E., Dritsou, V., Dialynas, E., Louis, C.: IDOMAL: the malaria ontology revisited. In: *Journal of Biomedical Semantics* (2013) 4-16
5. Shaban-Nejad, A., and Haarslev, V. Managing changes in distributed biomedical ontologies using hierarchical distributed graph transformation. *International Journal of Data Mining and Bioinformatics*, 11(1): 53-83 (2015)
6. Vanderwalk, B. P., McCarthy, E. L., Wilkinson, M. D. In: *SHARE: A Semantic Web Query Engine for Bioinformatics*. Springer Berlin Heidelberg, Berlin, Heidelberg (2009) 367-369
7. Al Manir, M. S., Riazanov, A, Boley, H., Klein, A., Baker C. J.: Valet SADI: Provisioning SADI web services for semantic querying of relational databases. In: *Proc. of IDEAS'16*, New York, NY, USA, ACM (2016) 248-255.
8. Gonçalves, R., Parsia, P., Sattler, U.: Ecco: a hybrid diff tool for OWL 2 ontologies. In: *Proc. of OWLED 2012*.
9. Liquibase: source control for your database – <http://www.liquibase.org>