

# A Semantic Framework to Improve Interoperability of Malaria Surveillance Systems

Jon Hael Simon Brenas<sup>\*1</sup>, Mohammad S. Al-Manir<sup>2</sup>, Kate Zinszer<sup>3</sup>, Christopher J. Baker<sup>2</sup> and Arash Shaban-Nejad<sup>1</sup>

<sup>1</sup>UTHSC/ORNL, Memphis, TN, USA; <sup>2</sup>University of New Brunswick, Saint John, NB, Canada; <sup>3</sup>Université de Montréal, Montréal, QC, Canada

## Objective

Malaria is one of the top causes of death in Africa and some other regions in the world. Data driven surveillance activities are essential for enabling the timely interventions to alleviate the impact of the disease and eventually eliminate malaria. Improving the interoperability of data sources through the use of shared semantics is a key consideration when designing surveillance systems, which must be robust in the face of dynamic changes to one or more components of a distributed infrastructure. Here we introduce a semantic framework to improve interoperability of malaria surveillance systems (SIEMA).

## Introduction

In 2015, there were 212 million new cases of malaria, and about 429,000 malaria death, worldwide. African countries accounted for almost 90% of global cases of malaria and 92% of malaria deaths. Currently, malaria data are scattered across different countries, laboratories, and organizations in different heterogeneous data formats and repositories. The diversity of access methodologies makes it difficult to retrieve relevant data in a timely manner. Moreover, lack of rich metadata limits the reusability of data and its integration. The current process of discovering, accessing and reusing the data is inefficient and error-prone profoundly hindering surveillance efforts.

As our knowledge about malaria and appropriate preventive measures becomes more comprehensive malaria data management systems, data collection standards, and data stewardship are certain to change regularly. Collectively these changes will make it more difficult to perform accurate data analytics or achieve reliable estimates of important metrics, such as infection rates. Consequently, there is a critical need to rapidly re-assess the integrity of data and knowledge infrastructures that experts depend on to support their surveillance tasks.

## Methods

In order to address the challenge of heterogeneity of malaria data sources we recruit domain specific ontologies in the field (e.g. IDOMAL (1)) that define a shared lexicon of concepts and relations. These ontologies are expressed in the standard Web Ontology Language (OWL).

To overcome challenges in accessing distributed data resources we have adopted the Semantic Automatic Discovery & Integration framework (SADI) (2) to ensure interoperability. SADI provides a way to describe services that provide access to data, detailing inputs and outputs of services and a functional description. Existing ontology terms are used when building SADI Service descriptions. The services can be discovered by querying a registry and combined into complex workflows. Users can issue SPARQL syntax to a query engine which can plan complex workflows to fetch actual data, without having to know how target data is structured or where it is located.

In order to tackle changes in target data sources, the ontologies or the service definitions, we create a Dashboard (3) that can report any changes. The Dashboard reuses some existing tools to perform a series of checks. These tools compare versions of ontologies and databases allowing the Dashboard to report these changes. Once a change has

been identified, as series of recommendations can be made, e.g. services can be retired or updated so that data access can continue.

## Results

We used the Mosquito Insecticide Resistance Ontology (MIRO) (5) to define the common lexicon for our data sources and queries. The sources we created are CSV files that use the IRbase (4) schema. With the data defined using we specified several SPARQL queries and the SADI services needed to answer them. These services were designed to enabled access to the data separated in different files using different formats. In order to showcase the capabilities of our Dashboard, we also modified parts of the service definitions, of the ontology and of the data sources. This allowed us to test our change detection capabilities. Once changes were detected, we manually updated the services to comply with a revised ontology and data sources and checked that the changes we proposed were yielding services that gave the right answers. In the future, we plan to make the updating of the services automatic.

## Conclusions

Being able to make the relevant information accessible to a surveillance expert in a seamless way is critical in tackling and ultimately curing malaria. In order to achieve this, we used existing ontologies and semantic web services to increase the interoperability of the various sources. The data as well as the ontologies being likely to change frequently, we also designed a tool allowing us to detect and identify the changes and to update the services so that the whole surveillance systems becomes more resilient.

## Keywords

Semantic; Interoperability; Malaria Surveillance

## Acknowledgments

This work is supported by the Bill and Melinda Gates Foundation.

## References

1. P. Topalis, E. Mitraka, V Dritsou, E. Dialynas and C. Louis, "IDOMAL: the malaria ontology revisited" in *Journal of Biomedical Semantics*, vol. 4, no. 1, p. 16, Sep 2013.
2. M. D. Wilkinson, B. Vandervalk and L. McCarthy, "The Semantic Automated Discovery and Integration (SADI) web service design-pattern, API and reference implementation" in *Journal of Biomedical Semantics*, vol. 2, no. 1, p. 8, 2011.
3. J.H. Brenas, M.S. Al-Manir, C.J.O. Baker and A. Shaban-Nejad, "Change management dashboard for the SIEMA global surveillance infrastructure", in *International Semantic Web Conference*, 2017
4. E. Dialynas, P. Topalis, J. Vontas and C. Louis, "MIRO and IRbase: IT Tools for the Epidemiological Monitoring of Insecticide Resistance in Mosquito Disease Vectors", in *PLOS Neglected Tropical Diseases* 2009

\*Jon Hael Simon Brenas

E-mail: jhael@uthsc.edu

