The SIEMA Platform: Improving DHIS2 Data Uses in Malaria Surveillance

Mohammad Sadnan Al Manir[1], Jon Haël Brenas[2], Kate Zinszer[3], Christopher J.O. Baker[1,4] and Arash Shaban-Nejad[2]

[1] Department of Computer Science
University of New Brunswick, Saint John, Canada
{jhael,ashabann}[at]uthsc.edu
[2] The University of Tennessee Health Science Center (UTHSC)
Oak Ridge National Laboratory (ORNL)
Center for Biomedical Informatics, Department of Pediatrics, Memphis, TN, USA
{sadnan.almanir,bakerc}[at]unb.ca
[3] École de santé publique, Département de médecine sociale et préventive, Université de Montréal, Canada
{kate.zinszer}[at]umontreal.ca
[4] IPSNP Computing Inc. Saint John, Canada

Abstract. The Semantics, Interoperability, and Evolution for Malaria Analytics (SIEMA) platform a provides solution to integration and interoperability challenges associated with distributed malaria data. SIEMA leverages (i) community developed ontologies for facilitating standardization, (ii) SADI Semantic Web services for discovering and accessing distributed data, (iii) an analytics dashboard for detecting mission-critical changes interrupting the system interoperability, (iv) Valet SADI tool for building SADI services and re-actively rebuilding services to accommodate the changes in terminologies, and (v) the HYDRA query engine that provides the functionality to create SPARQL queries graphically and perform the automatic discovery, invocation, orchestration, and execution of services. Whereas the open-source surveillance software District Health Information Software 2 (DHIS2) provides improved malaria surveillance in many countries, SIEMA goes beyond the capabilities of such systems by addressing the needs for semantics, greater interoperability and evolution or change management needs of dynamic surveillance systems.

1 Introduction

Malaria is an infectious disease caused by parasitic micro-organisms of the Plasmodium species and transmitted by mosquitoes of the Anopheles Gambiae species. Malaria is endemic in 91 countries and responsible for an estimated 216 million cases in 2016 [1]. Efficient sharing of data across countries, between languages, scientists and medical professionals is a major obstacle to reducing the burden of malaria. In order to store, share, and access malaria data dedicated platforms capable of unifying these tasks and operations have emerged. An example of such a

platform is DHIS2[5] that is currently being used in more than 60 countries. DHIS2 contributes to the integration of target data and syntactic interoperability between systems by using standard data formats, bespoke interfaces, codes and terminologies. DHIS2 however, falls short of providing semantic interoperability due to the lack of meanings associated to the archived data. Moreover, it lacks support for problems associated with change management where changes in the resources and terminologies result in the failure of the surveillance operations. In this work, we demonstrate the utility of an alternative approach implemented within the SIEMA platform [2,3] where access to data is provided using SADI Semantic Web services [4]. Semantic interoperability of services is achieved through the use of standardized terminologies and ontologies. SIEMA also supports the preservation of semantic interoperability by rebuilding services to address the unexpected changes in resources responsible for the service-downtime.

2 Demonstration of SIEMA features

In order to illustrate the extended capabilities of the SIEMA platform data, resources from DHIS2 have been selected as the source data in this demo. DHIS2 resources contain, for example, information about the criteria for "Immunization" of children which are distributed in separate tables. We have created and deployed a number of SADI Semantic Web services that access source data allowing an end user to ask questions about immunization such as "Which charts contain information about a Fully Immunized child?". Although this question seems fairly simple, for the average user composing such a query involves accessing three tables and requires knowledge of SQL. A SPARQL query may be easier to create for the user with the knowledge of the terminologies, but the raw SPARQL syntax is still challenging to learn. We demonstrate that the GUI of HYDRA[6] query engine allows a user to compose such a query with relative ease using the keyword-based graphical query canvas. HYDRA also performs the automatic translation of the query graph into the equivalent raw SPARQL syntax. The query is subsequently executed discovering, invoking, and orchestrating the services automatically from the query engine to return immunization data.

In addition, we also demonstrate that when a change occurs in a target resource, for example if a term used in a SADI service description is renamed or added, the service may no longer be able to retrieve the target data and the query fails. A dashboard [5] implemented within SIEMA can detect these changes, identify their types, and report them to the user. It also shows the status of services and pinpoints which services failed to operate due to the changes. In turn, the Valet SADI tool [6] which is used to generate services automatically is then triggered to rebuild unresponsive services. This minimizes the service's downtime, and consequently queries can again be executed. Thus, problems originating from a change are addressed and the semantic interoperability of the system is preserved.

[5] `http://dhis2.org/`
[6] `http://ipsnp.com/hydra/`

References

1. WHO. *World Malaria Report 2017*. Geneva, Switzerland, 2016.
2. Mohammad Sadnan Al Manir, Jon Haël Brenas, Christopher JO Baker, and Arash Shaban-Nejad. A surveillance infrastructure for malaria analytics: Provisioning data access and preservation of interoperability. *JMIR Public Health Surveill*, 4(2):e10218, Jun 2018.
3. Jon Haël Brenas, Mohammad Sadnan Al Manir, Christopher J. O. Baker, and Arash Shaban-Nejad. A malaria analytics framework to support evolution and interoperability of global health surveillance systems. *IEEE Access*, 5:21605–21619, 2017.
4. Mark Wilkinson, Benjamin Vandervalk, and Luke McCarthy. The Semantic Automated Discovery and Integration (SADI) Web service Design-Pattern, API and Reference Implementation. *Journal of Biomedical Semantics*, 2(1):8, 2011.
5. Jon Haël Brenas, Mohammad Sadnan Al Manir, Christopher J. O. Baker, and Arash Shaban-Nejad. A change management dashboard for the SIEMA global malaria surveillance infrastructure. In *Proceedings of the ISWC 2017 Posters & Demonstrations and Industry Tracks co-located with 16th International Semantic Web Conference (ISWC 2017), Vienna, Austria, October 23rd - to - 25th, 2017.*, 2017.
6. Mohammad Sadnan Al Manir, Alexandre Riazanov, Harold Boley, Artjom Klein, and Christopher J. O. Baker. Valet SADI: provisioning SADI web services for semantic querying of relational databases. In *Proceedings of the 20th International Database Engineering & Applications Symposium, IDEAS 2016, Montreal, QC, Canada, July 11-13, 2016*, pages 248–255, 2016.