

Applying a Mapping Quality Framework in Cloud Native Monitoring

Alex Randles¹, Declan O’Sullivan¹, John Keeney² and Liam Fallon²

¹ ADAPT Centre for Digital Content, Trinity College Dublin, Ireland

² Ericsson Software Technology, Athlone, Co. Westmeath, Ireland

Abstract

The linked data generation process is a complex process which involves multiple stakeholders in the definition of mapping artefacts which transform source data into linked data representation. The creation of these mapping artefacts is error prone, and the quality of the artefacts should be assessed prior to the generation of the linked data. Producing high quality mappings will result in high quality linked data and provide higher confidence to linked data consumers. Furthermore, the source data of these mappings should be regularly monitored to detect data changes which could impact the quality of the resulting linked data dataset. A process designed to offer fresh and high-quality data will benefit both data consumers and producers. Applying the process to a real-world use case demonstrates the feasibility and effectiveness of the process. In this paper we describe a process designed to improve quality within the linked data generation process. Furthermore, we describe the application of the process to a real-world cloud monitoring use case

Keywords

Cloud native monitoring, Linked data, Data quality.

1. Introduction

Cloud monitoring data typically contains time series information with the metric information stored with a corresponding timestamp. However, commonly used cloud monitoring platforms such as Prometheus [1] do not as yet offer the ability to easily avail of machine-readable data which can limit the usability and utility of such data for analysis. Converting the monitoring data into an analysis-ready machine-readable format (such as the W3C linked data representation³) will enable additional knowledge discovery.

Transforming the time series data collected from a system such as Prometheus [1] into linked data format requires the definition of the mapping artefacts which define the rules for transformation. Creating these mapping artefacts can be a complex time-consuming task, which is frequently error prone [2]. It is argued that dealing with the quality issues for these mapping artefacts themselves will ensure that data quality issues do not propagate into the resulting dataset [3]. The MQV framework [4,5] which is designed to assess and refine the quality of these mappings can be used for the purpose of producing high quality mappings. Details of the core assessment and refinement aspects of the MQV framework is being presented in a paper in the research and innovation track of the SEMANTiCS 2022 conference. The framework has recently been extended to include a change detection component. The component is designed to ensure the freshness of linked data generated through the mapping artefacts by detecting and analyzing source data changes. The framework enables the impact of these detected changes to be assessed, and necessary action can be taken to ensure the freshness of the associated linked data for consumers.

SEMANTICS 2022 EU: 18th International Conference on Semantic Systems, September 13-15, 2022, Vienna, Austria

EMAIL: alex.randles@adaptcentre.ie (A. 1); declan.osullivan@adaptcentre.ie (A. 2); john.keeney@est.tech (A. 3); liam.fallon@est.tech (A.4)



© 2022 Copyright for this paper by its authors.

Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings (CEUR-WS.org)

³ W3C linked data representation at <https://www.w3.org/standards/semanticweb/data>

In this paper we provide a discussion of the evolution of the MQV framework and the application to a real-world cloud monitoring use case. The paper is structured as follows: Section 2 describes the background & motivation of this work; Section 3 discusses the application of the framework to the cloud monitoring use case; Section 4 describes related work and Section 5 concludes the paper and describes future work.

2. Background & Motivation

Prometheus [1] is a widely used open-source system designed for cloud native monitoring. Prometheus collects and stores its metrics as time series data. The main component of the Prometheus ecosystem is the Prometheus server which scrapes time series data from target servers. The Prometheus API allows queries to be executed which return metric information. However, the API cannot return data in a graph format which limits the analysis, linkability and expressiveness of the data.

In our research we propose to transform the Prometheus time series data into a more analysis and linkable friendly format, namely linked data. The benefits of transformation include easier data processing and interlinking with semantically similar data. Linking the data with other monitoring data could be used to drive an intent-based system, which can define and accomplish required goals automatically. Within the state-of-the-art limited work could be found which tackles the challenge of transforming Prometheus time series data into a linkable graph format. It is hoped the approach will provide greater insight and flexibility when performing cloud native monitoring.

3. Use case

A framework named the Prometheus RDF Generator has been designed to transform statistical data from Prometheus [1] into linked data format. **Figure 1** shows a screenshot of the current implementation of the framework.

Connect to a Prometheus server and generate RDF network data.

Prometheus Server IP address ⓘ

Predefined Metric selection ⓘ

RDF data successfully generated!

Prometheus API query details
 Metric: HTTP Code Responses
 API Query used: 'prometheus_http_requests_total'

Prometheus API Response
 The table below has been generated by querying the [RDF](#) which was created from the Prometheus API response using an R2RML [Mapping](#).

Metric						
ID	hasInstance	hasValue	hasHandler	hasName	hasCode	hasJob
0	localhost:9090	8	/metrics	prometheus_http_requests_total	200	prometheus

Figure 1: Screenshot of the Prometheus metric selection (left) and sample results generated (right)

The process starts with the users entering the IP address of the Prometheus server and selecting a metric from a predefined list, which is shown on the left side of the figure. Thereafter, an API request is made to the server which will return a response, which is shown on the right side of the figure. The response is stored in a relational database which allows the data to be converted into linked data format using R2RML [6] mappings. The mapping uses an ontology currently under development which has been designed specifically to model the time series metric data. The result is expressive machine-readable information which can be easily queried and interlinked with other related information.

The MQV framework [4,5] is designed to assess and refine the quality of mappings and its main functionality is described in a paper in the research and innovation track of the SEMANTiCS 2022 conference. The paper describes the framework design and a user evaluation which was conducted with a sample size of 58 participants. Improvements have been made to the framework based on the evaluation such as improved interface aesthetics. Furthermore, in the time since the submission of the SEMANTiCS paper, the framework has been extended to include a change detection component and applied as part of an internship with Ericsson Software Technology.

The component is designed to ensure the freshness of the linked data by detecting and analyzing the changes which occur in the source data. The component regularly monitors the source data to detect new changes and uplifts them into linked data, represented using the Change Detection Ontology (CDO)⁴. The ontology has been specifically designed by us to represent source data changes. Moreover, the component allows users to define a notification policy, which defines when users will be notified of changes. The policy is defined by users entering thresholds into the user interface. These thresholds define how many changes or types of changes should be detected before a notification is sent. The user interface input is uplifted into linked data and represented using Rei policy ontology [7]. Mappings associated with the source data can be uploaded to the component, which can be analyzed to detect the impact of these changes on them and the resulting linked data. For example, if data has been added to the source data of the mapping, the dataset should be regenerated to capture the new information and ensure the freshness of the data. **Figure 2** shows a screenshot of the newer version of the framework.

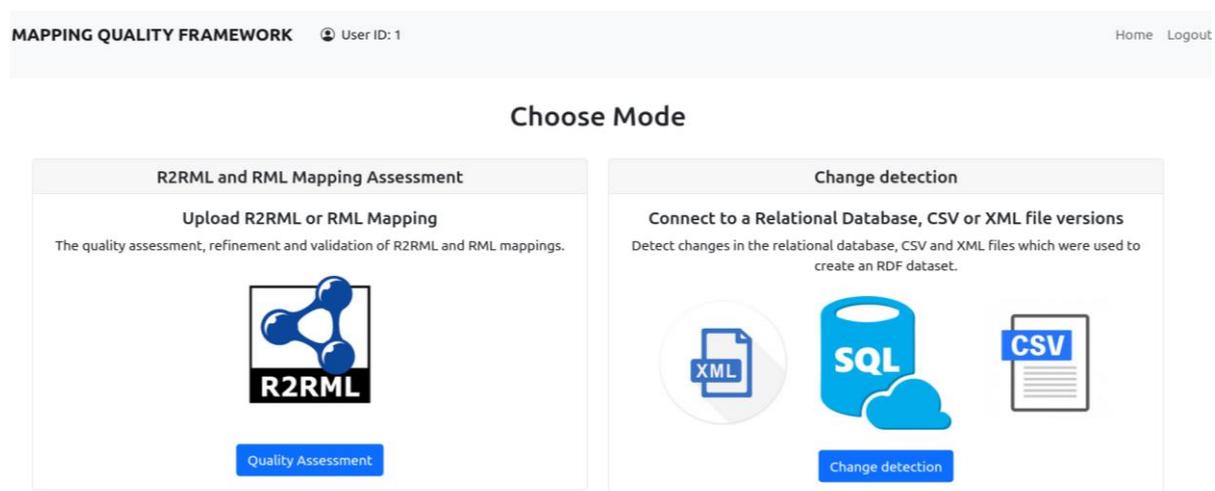


Figure 2: Screenshot of the Mapping Assessment (left) and the Change Detection (right) mode choice

The framework allows users to select two modes of operation. These modes include i) **Mapping Assessment** and ii) **Change Detection**. These two modes have been used during the design of the Prometheus RDF Generator system and their usage is outlined below.

Mapping Assessment. The mapping assessment component was used to assess the quality of mappings⁵ which were used to uplift the monitoring information. The local ontology functionality which was previously described in [4,5] was utilized as the ontology used to represent the information is still under development. The mappings were uploaded to the framework and a validation report⁶ was generated, which contains quality assessment and refinement information. The report generated shows that no quality issues have been detected within the mappings. Therefore, no quality refinement information is contained in the report.

Change Detection. The component is currently being used to consistently analyze the monitoring data which is being collected from the Prometheus server [1]. The data is stored within a relational

⁴ Change Detection Ontology specification at <https://alex-randles.github.io/Change-Detection-Ontology/>

⁵ R2RML mappings used to uplift Prometheus data at https://github.com/alex-randles/Semantics-2022-Demo-Paper/blob/main/prometheus_mapping.ttl

⁶ Validation report generated by framework at https://github.com/alex-randles/Semantics-2022-Demo-Paper/blob/main/validation_report.ttl

database by the Prometheus RDF Generator which the change detection component can monitor in order to detect new changes. Other source data formats can be exposed through a URL. Furthermore, the mappings used to uplift the data into linked data format have been uploaded, which allows the framework to assess the impact of changes on the mappings and resulting linked data dataset. Moreover, a notification policy has been defined to ensure we are notified of the changes when necessary. An example of a change detected when applied to the use case relates to the addition of a node exporter to the Prometheus server. A node exporter is software designed to collect metric information from a target server through exporting server and OS level information [1]. The addition of the node exporter to the Prometheus environment was detected by the change detection component and represented in graph format. **Figure 3** shows the components involved in the detection.

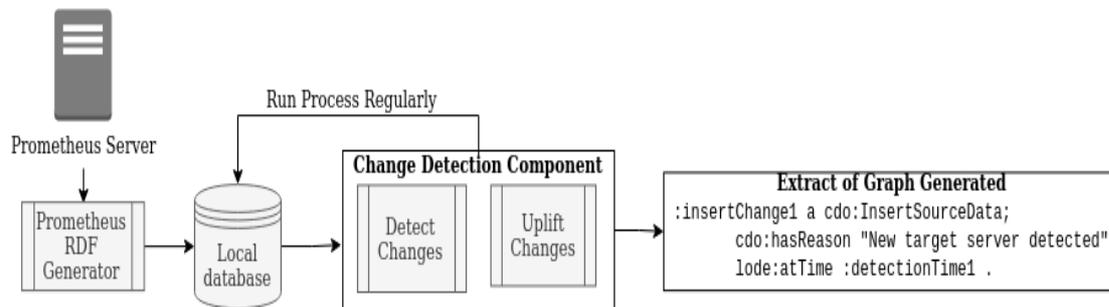


Figure 3: Integration of Change Detection Component with the Prometheus RDF Generator

The data from the Prometheus server is fed into the Prometheus RDF Generator, where it is processed and stored in a relational database before being uplifted into linked data format. The change detection component is connected to the database and regularly queries it to detect new changes. These changes are transformed into linked data format. The extract of the graph shows where the system has detected the new target server (:insertChange1) and the time it was detected (:detectionTime1).

4. Related work

The state of the art in mapping quality frameworks and dataset change frameworks has been reviewed.

Mapping Quality Frameworks. Resglass [8] is a rule driven approach which targets RML [9] mappings to detect inconsistencies using rules used to generate linked data datasets. The approach [3] provides a test driven approach which extends an existing RDF test case based architecture to assess and refine the quality of RML mappings. The approach [2] extends an existing quality assessment framework to target R2RML [6] mappings using metrics commonly used to assess the quality of the resulting dataset. Furthermore, the framework is extensible allowing additional metrics to be added.

Dataset Change Frameworks. DSNotify [10] is an approach designed to address change detection and link maintenance in linked data datasets. The approach includes a monitor component which regularly analyses the dataset for resource changes. sparqlPuSH [11] is a framework which enables proactive notification of updates within linked datasets. Users can subscribe to a subset of the dataset using a SPARQL query and will be notified when content within the subset changes. DELTA-LD [12] is designed to address broken links and synchronization of a linked dataset by identifying the delta between two versions of dataset. The approach detects changes by generating features from the properties and objects within the dataset which can be used for comparison.

5. Future work & Conclusion

Future work includes the completion of the development and deployment of the Prometheus RDF Generator. Furthermore, the ontology used to model the time series data will be evaluated using methods such as ontology competency questions [13]. Therefore, ensuring the ontology conforms to the design requirements. A second user evaluation will be completed on the MQV framework which will evaluate the additional change detection component functionality. The evaluation will involve users

interacting with the system using predefined tasks and standardized methods such as the Post-Study System Usability Questionnaire [14]. Furthermore, the ontology used within the change detection component to represent change information will be evaluated in a similar manner to the aforementioned methods.

Deploying the framework into a real-world use case is beneficial to demonstrate the usability and effectiveness of the design. The framework ensures that the mappings used within the Prometheus RDF Generator design are of the good quality, thus ensuring high quality data to the consumers. Furthermore, the change detection component allows changes within the source data to be detected and analyzed allowing the impact of changes on the mappings and resulting dataset to be assessed, therefore, promoting fresh linked data. It is hoped the combination of these two components within the use case will aid the data transformation process while demonstrating the feasibility of the design for others.

6. Acknowledgements

This research was conducted with the financial support of the SFI AI Centre for Research Training under Grant Agreement No. 18/CRT/6223 at the ADAPT SFI Research Centre at Trinity College Dublin. The ADAPT SFI Centre for Digital Media Technology is funded by Science Foundation Ireland through the SFI Research Centres Programme and is co-funded under the European Regional Development Fund (ERDF) through Grant # 13/RC/2106. Application of the MQV framework in the Prometheus use case has been undertaken as part of an internship in Ericsson Software Technology.

7. References

- [1] Prometheus [Internet]. Available from: <https://prometheus.io/>
- [2] Junior AC, Debattista J, O’Sullivan D. Assessing the Quality of R2RML Mappings. In: Joint Proceedings of the International Workshop On Semantics For Transport and on Approaches for Making Data Interoperable co-located with 15th Semantics Conference, Karlsruhe, Germany. CEUR-WS; 2019. (CEUR Workshop Proceedings; vol. 2447).
- [3] Dimou A, Kontokostas D, Freudenberg M, Verborgh R, Lehmann J, Mannens E, et al. Assessing and refining mappings to RDF to improve dataset quality. In: Lecture Notes in Computer Science. Springer Verlag; 2015. p. 133–49.
- [4] Randles A, O’Sullivan D. Assessing quality of R2RML mappings for OSI’s Linked Open Data portal. 4th Int Work Geospatial Linked Data ESWC 2021. 2021;
- [5] Randles A, O’Sullivan D. Evaluating Quality Improvement techniques within the Linked Data Generation Process. In: 18th International Conference on Semantics Systems. 2022.
- [6] Das S, Sundara S, Cyganiak R. R2RML: RDB to RDF Mapping Language. W3C Recomm [Internet]. 2012; Available from: <http://www.w3.org/TR/r2rml/>
- [7] Kagal L, others. Rei: A policy language for the me-centric project. 2002;
- [8] Heyvaert P, De Meester B, Dimou A, Verborgh R. Rule-driven inconsistency resolution for knowledge graph generation rules. *Semant Web*. 2019;10(6).
- [9] Dimou A, Vander Sande M, Colpaert P, Verborgh R, Mannens E, de Walle R. RML: a generic language for integrated RDF mappings of heterogeneous data. In: *Ldow*. 2014.
- [10] Popitsch N, Haslhofer B. DSNotify - A solution for event detection and link maintenance in dynamic datasets. *J Web Semant*. 2011 Sep;9(3):266–83.
- [11] Passant A, Mendes PN. sparqlPuSH: Proactive notification of data updates in RDF stores using PubSubHubbub [Internet]. Available from: <http://www.ldodds.com/blog/2010/04/rdf-dataset-notifications/>
- [12] Singh A, Brennan R, O’Sullivan D. DELTA-LD: A Change Detection Approach for Linked Datasets.
- [13] Bezerra C, Freitas F, Santana F. Evaluating ontologies with Competency Questions. In: Proceedings - 2013 IEEE/WIC/ACM International Joint Conference on Web Intelligence and Intelligent Agent Technology - Workshops, WI-IATW 2013. 2013.
- [14] Lewis JR. Psychometric Evaluation of the PSSUQ Using Data from Five Years of Usability Studies. *Int J Hum Comput Interact*. 2002 Sep;14(3–4):463–88.