A Declarative Rule-Based Formalization of R2RML: Bridging Knowledge Graph Construction and Database Theory

Ali Elhalawati [1,†], Anastasia Dimou [1] , and Jan Van den Bussche [2]

[1] KU Leuven - Flanders Make@KULeuven - Leuven.AI, Leuven, Belgium
{ali.elhalawati,anastasia.dimou}@kuleuven.be
[2] Hasselt University, Hasselt, Belgium
jan.vandenbussche@uhasselt.be
† Presenter

Declarative mapping languages are widely used to generate knowledge graphs from heterogeneous data sources [7]. Among them, R2RML [2] is a widely adopted W3C-recommendation for generating knowledge graphs, in the form of RDF graphs, from relational databases. R2RML enables viewing relational data as RDF, using a structure and vocabulary defined by the mapping author. Several works [4, 5, 6] have attempted to formalize the semantics of R2RML; however, none has provided a complete and precise formalization. As a result, subsequent mapping languages [7] extended R2RML without clearly distinguishing which components were reused or were newly introduced.

In this presentation, we show how to define declarative, computable, rule-based formalizations of R2RML as we demonstrated in [3]. For our formalizations we use Datalog [1], a well-established rule-based query language in database theory widely used for declarative data processing and reasoning in deductive databases. We present two formalization approaches with different goals. The first translates a given set of R2RML mappings and input data into a customized Datalog program that can be efficiently executed to generate RDF graphs, with triples and quadruples represented as relations. The second is a "universal", fixed Datalog program that can be executed on any R2RML mapping and input data, eliminating the need for translating R2RML mappings into Datalog rules, which offers greater expressivity at the cost of reduced efficiency compared to the first approach. Finally, we demonstrate how our Datalog-based formalization can be adapted to generate RML mappings, enabling the construction of RDF graphs from heterogeneous data sources beyond relational databases.

We validated our Datalog-based semantics by executing the R2RML test cases with a prototype implementation (`https://github.com/dtai-kg/R2RML2Datalog-Tests`). Furthermore, we show that, even without optimization, our prototype (`https://github.com/dtai-kg/R2RML2Datalog-Translator`) performs competitively with leading R2RML systems on a widely used benchmark covering diverse R2RML mapping scenarios.

Our work lays the foundation for formally studying R2RML and its extensions, unlocks the benefits of Datalog reasoning in RDF generation, and introduces an efficient approach for producing RDF graphs using standard Datalog engines.

References

[1] Serge Abiteboul, Richard Hull, and Victor Vianu, editors. *Foundations of Databases: The Logical Level*. Addison-Wesley Longman Publishing Co., Inc., 1st edition, 1995. URL:

http://webdam.inria.fr/Alice/.

[2] Souripriya Das, Seema Sundara, and Richard Cyganiak. R2RML: RDB to RDF Mapping Language. Working group recommendation, 2012. URL: http://www.w3.org/TR/r2rml/.

[3] Ali Elhalawati, Anastasia Dimou, and Jan Van den Bussche. A Declarative Formalization of R2RML Using Datalog and Its Efficient Execution. In *RuleChallenge@RuleML+RR 2025, Istanbul, Turkey.*, 2025.

[4] Enrique Iglesias, Samaneh Jozashoori, and Maria-Esther Vidal. Scaling up knowledge graph creation to large and heterogeneous data sources. *Journal of Web Semantics*, 75, 2023. doi:10.1016/j.websem.2022.100755.

[5] Mariano Rodriguez-Muro and Martín Rezk. Efficient SPARQL-to-SQL with R2RML mappings. *Journal of Web Semantics*, 33:141–169, 2015.

[6] Juan F. Sequeda. On the semantics of R2RML and its relationship with the direct mapping. In *Proceedings of the 12th International Semantic Web Conference Posters & Demonstrations Track (ISWC-PD '13)*, volume 1035, page 193–196. CEUR-WS.org, 2013. URL: https://ceur-ws.org/Vol-1035/iswc2013_poster_4.pdf.

[7] Dylan Van Assche, Thomas Delva, Gerald Haesendonck, Pieter Heyvaert, Ben De Meester, and Anastasia Dimou. Declarative RDF graph generation from heterogeneous (semi-) structured data: A systematic literature review. *Journal of Web Semantics*, 75, 2023. doi:10.1016/j.websem.2022.100753.