Preface for the 5th Edition of the International Knowledge Graph Construction Workshop

David Chaves-Fraga[1,2], Anastasia Dimou[2,3,4], Ana Iglesias-Molina[5], Umutcan Serles[6] and Dylan Van Assche[7]

[1]*Universidade de Santiago de Compostela, Departamento de Electrónica e Computación, Santiago de Compostela, Spain*
[2]*KU Leuven, Department of Computer Science, Sint-Katelijne-Waver, Belgium*
[3]*Flanders Make – DTAI-FET*
[4]*Leuven.AI – KU Leuven institute for AI, B-3000 Leuven, Belgium*
[5]*Universidad Politécnica de Madrid, Campus de Montegancedo, Boadilla del Monte, Spain*
[6]*Semantic Technology Institute Innsbruck, Universität Innsbruck, Austria*
[7]*IDLab, Dept of Electronics and Information Systems, Ghent University – imec, Belgium*

More and more knowledge graphs are constructed for private use, e.g., the Amazon Product Graph [1] or the Fashion Knowledge Graph by Zalando[1],or public use, e.g., DBpedia[2] or Wikidata[3]. While techniques to automatically construct KGs from existing Web objects exist (e.g., scraping Web tables), there is still room for improvement. So far, constructing knowledge graphs was considered an engineering task, however, more scientifically robust methods keep on emerging. These methods were widely questioned for their verbosity, low performance or difficulty of use, while the data sources' variety and complexity cause further syntax and semantic interoperability issues.

Declarative methods (mapping languages) for describing rules to construct knowledge graphs and approaches to execute those rules keep on emerging. Nevertheless constructing knowledge graphs is still not a straightforward task because several existing challenges remain and yet the barriers to construct knowledge graphs are not lowered enough to be easily and broadly adopted by industry. These reasons and the vastly populated knowledge graph construction W3C Community Group[4] show that there are still open questions that require further investigation to come up with groundbreaking solutions.

Addressing challenges related to knowledge graphs construction requires well-founded research, including the investigation of concepts and development of tools as well as methods for their evaluation. R2RML was recommended in 2012 by W3C, and since then, different extensions, alternatives and implementations were proposed [2, 3, 4]. Certain approaches followed the ETL-like paradigm, e.g., SDM-RDFizer [5], RocketRML [6], and FunMap [7], while

Fifth International Workshop On Knowledge Graph Construction Co-located with the ESWC 2024, 27th May 2024, Crete, Greece
david.chaves@upm.es (D. Chaves-Fraga); anastasia.dimou@kuleuven.be (A. Dimou); ana.iglesiasm@upm.es (A. Iglesias-Molina); umutcan.serles@sti2.at (U. Serles); dylan.van.assche@ugent.be (D. V. Assche)

[1]https://engineering.zalando.com/posts/2018/03/semantic-web-technologies.html
[2]https://www.dbpedia.org/resources/knowledge-graphs/
[3]https://www.wikidata.org/wiki/Wikidata:Main_Page
[4]http://w3.org/community/kg-construct

others the query-answering paradigm, e.g., Ultrawrap [8], Morph-RDB [9] and Ontop [10]. Besides R2RML-based extensions, alternatives were proposed, e.g., SPARQL-Generate [11] and ShExML [12], as well as methods to perform data transformations while constructing knowledge graphs, e.g., FnO [13] and FunUL [14].

The fifth edition of the knowledge graph construction workshop[5] has a special focus on time on novel techniques, frameworks, architectures, and tools for the new extensions of RML such as RDF Collections and Containers, and RDF-Star support and the 2023 release of the RDF Mapping Language (RML) [15] in general. It also included:

- *Keynote.* The workshop includes the keynote from Lionel Tailhardat (Orange): "Anomaly Detection For Telco Companies: Challenges And Opportunities In Knowledge Graph Construction"
- *The Second Knowledge Graph Construction Challenge.* The edition of this year's challenge has a double objective: benchmarking systems to (i) find which RDF graph construction system optimizes for metrics i.e. execution time, CPU and memory usage; and (ii) how compliant are they with the 2023 revision of RML and its new modules.

The final goal of the event is to provide a venue for scientific discourse, systematic analysis and rigorous evaluation of languages, techniques and tools, as well as practical and applied experiences and lessons-learned for constructing knowledge graphs from academia and industry.

Eight papers were submitted. The reviews were open and public, and hosted at Open Review[6]. Each paper received at least three reviews from reviewers with different background and status. Each paper received a review from a senior, a junior and an industry researcher.

Five papers were accepted and one was conditionally accepted. Five of the accepted papers were long papers and one was a short paper. The following papers were accepted for publication and presented at the workshop:

- Not Everybody Speaks RDF: Knowledge Conversion between Different Data Representations [16].
- BURPing Through RML Test Cases [17].
- Propagating Ontology Changes to Declarative Mappings in Construction of Knowledge Graphs [18].
- RML-view-to-CSV: A Proof-of-Concept Implementation for RML Logical Views [19].
- R2[RML]-ChatGPT Framework [20].
- Towards Self-Configuring Knowledge Graph Construction Pipelines using LLMs - A Case Study with RML [21].

During the workshop, the second edition of the Knowledge Graph Construction Challenge was organized with two different tracks: (i) conformance with the new RML modules, and (ii) performance of engines on the same hardware.

The first track around conformance with the new RML modules encouraged developers of RML engines to support the specifications of the new RML modules by evaluating their engines

[5] http://w3id.org/kg-construct/workshop/2024
[6] https://openreview.net/group?id=eswc-conferences.org/ESWC/2024/Workshop/KGCW

against 365 test cases provided by the maintainers of each RML module. RML-Core (238 test cases), which focus on the core parts of RDF generation, provides the biggest number of test cases, followed by RML-IO (67 test cases) to access various data sources and targets. Data transformations with FnO were also present through the RML-FNML (13 test cases) module. Newer modules e.g. RML-Star (18 test cases) for RDF-Star support and RML-CC (29 test cases) to generate RDFS Collections & Containers provided new challenges for existing engines as they impact the RDF generation process. We had 5 participating engines for the first track: RMLMapper [2], SDM-RDFizer [5], mapping-template [16], RPT/SANSA [22], and BURP [17].

The second track around performance was similar to the previous edition except that now each participant had access to a common hardware environment. This way, each engine had the same restrictions regarding CPU and RAM. Through this track, we wanted to not only focus on execution time but also resource consumption of each engine. This track consisted of 2 parts: (i) artificial data for analyzing specific parameters of the construction process e.g. joins, data size, mappings, and (ii) real-life data of the GTFS Madrid Benchmark to evaluate approaches in real use cases. We had 6 participating engines for the second track: mapping-template [16], FlexRML [23], RMLWeaver-js [24], RPT/Sansa [22], RMLStreamer [25], and RML-view-to-CSV+RMLStreamer [19].

Several participants also submitted a report of their participation in one or both tracks. The following reports are included in the proceedings:

- RMLStreamer supported by RML-view-to-CSV in the Performance Track of the KGCW Challenge 2024 [26].
- RMLWeaver-JS: An Algebraic Mapping Engine in the KGCW Challenge 2024 [24].
- Performance Results of FlexRML in the KGCW Challenge 2024 [27].
- Backwards or Forwards? [R2]RML Backwards Compatibility in RMLMapper [28].
- The Conformance of an RML Processor Built from Scratch to Validate RML Specifications and Test Cases [29].
- Results for Knowledge Graph Creation Challenge 2024: SDM-RDFizer [30].
- KGCW2024 Challenge Report: RDFProcessingToolkit [31].

Organizing Committee

- David Chaves-Fraga, Universidade de Santiago de Compostela
- Anastasia Dimou, KU Leuven, Flanders Make, Leuven.AI
- Dylan Van Assche, Ghent University – imec – IDLab
- Ana Iglesias-Molina, Universidad Politécnica de Madrid
- Umutcan Serles, University of Innsbruck

Program Committee

- Anelia Kurteva, Delft University of Technology
- Beatriz Esteves, Universidad Politécnica de Madrid

- Ben De Meester, Ghent University – imec – IDLab
- Bram Steenwinckel, Ghent University – imec – IDLab
- Christophe Debruyne, Liège University
- Claus Stadler, University of Leipzig
- Davide Lanti, Free University of Bozen
- Edna Ruckhaus Magnus, Universidad Politécnica de Madrid
- Els de Vleeschauwer, Ghent University
- Enrique Antonio Iglesias, Leibniz University of Hannover
- Ernesto Jimenez-Ruiz, City, University of London
- Femke Ongenae, Ghent University
- Franck Michel, CNRS
- Gertjan De Mulder, Ghent University – imec – IDLab
- Giorgos Flouris, FORTH-ICS
- Hannes Voigt, TU Dresden
- Herminio García-González, Kazerne Dossin
- Ibai Guillén-pacho, Universidad Politécnica de Madrid
- Ioannis Dasoulas, KU Leuven
- Jakub Klímek, Charles University
- Juliette Opdenplatz, Universität Innsbruck
- Jürgen Umbrich, Vienna University of Economics and Business
- Manolis Koubarakis, National and Kapodistrian University of Athens
- Maria-Esther Vidal, Leibniz University of Hannover
- Mario Scrocca, Cefriel
- Markus Schröder, German Research Center for AI
- Michael Freund, Fraunhofer
- Oscar Corcho, Universidad Politécnica de Madrid
- Pano Maria, Skemu
- Samaneh Jozashoori, metaphacts GmbH
- Sergio José Rodríguez Méndez, Australian National University
- Sitt Min Oo, Ghent University – imec – IDLab
- Sven Lieber, Royal Library Of Belgium
- Tobias Schweizer, SWITCH
- Vladimir Alexiev, Ontotext

References

[1] X. L. Dong, X. He, A. Kan, X. Li, Y. Liang, J. Ma, Y. E. Xu, C. Zhang, T. Zhao, G. Blanco Saldana, S. Deshpande, A. Michetti Manduca, J. Ren, S. P. Singh, F. Xiao, H.-S. Chang, G. Karamanolakis, Y. Mao, Y. Wang, C. Faloutsos, A. McCallum, J. Han, AutoKnow: Self-Driving Knowledge Collection for Products of Thousands of Types, KDD '20, Association for Computing Machinery, New York, NY, USA, 2020, p. 2724–2734.

[2] A. Dimou, M. V. Sande, P. Colpaert, R. Verborgh, E. Mannens, R. V. de Walle, RML: A Generic Language for Integrated RDF Mappings of Heterogeneous Data, in: Proceedings of the 7th Workshop on Linked Data on the Web (LDOW), 2014.
[3] D. Chaves-Fraga, F. Priyatna, I. Perez-Santana, O. Corcho, Virtual Statistics Knowledge Graph Generation from CSV files, in: Emerging Topics in Semantic Technologies: ISWC 2018 Satellite Events, Studies on the Semantic Web, IOS Press, 2018.
[4] F. Michel, L. Djimenou, C. Faron-Zucker, J. Montagnat, xR2RML: Relational and Non-Relational Databases to RDF Mapping Language, Technical Report, 2017.
[5] E. Iglesias, S. Jozashoori, D. Chaves-Fraga, D. Collarana, M.-E. Vidal, SDM-RDFizer: An RML Interpreter for the Efficient Creation of RDF Knowledge Graphs, in: Proceedings of the 29th ACM International Conference on Information & Knowledge Management, 2020, pp. 3039–3046.
[6] U. Şimşek, E. Kärle, D. Fensel, RocketRML - A NodeJS implementation of a Use-Case Specific RML Mapper, in: Proceedings of the 1st Workshop on Knowledge Graph Building, 2019.
[7] S. Jozashoori, D. Chaves-Fraga, E. Iglesias, M.-E. Vidal, O. Corcho, FunMap: Efficient Execution of Functional Mappings for Knowledge Graph Creation, in: International Semantic Web Conference, Springer, 2020, pp. 276–293.
[8] J. F. Sequeda, D. P. Miranker, Ultrawrap: SPARQL execution on relational data, Web Semantics: Science, Services and Agents on the WWW (2013).
[9] F. Priyatna, O. Corcho, J. Sequeda, Formalisation and Experiences of R2RML-based SPARQL to SQL Query Translation Using Morph, in: Proceedings of the 23rd International Conference on World Wide Web, 2014.
[10] D. Calvanese, B. Cogrel, S. Komla-Ebri, R. Kontchakov, D. Lanti, M. Rezk, M. Rodriguez-Muro, G. Xiao, Ontop: Answering SPARQL Queries over Relational Databases, Semantic Web Journal (2017).
[11] M. Lefrançois, A. Zimmermann, N. Bakerally, A SPARQL Extension for Generating RDF from Heterogeneous Formats, in: The Semantic Web: 14th International Conference, 2017.
[12] H. García-González, I. Boneva, S. Staworko, J. E. Labra-Gayo, J. M. C. Lovelle, ShExML: improving the usability of heterogeneous data mapping languages for first-time users, PeerJ Computer Science 6 (2020) e318.
[13] B. De Meester, A. Dimou, R. Verborgh, E. Mannens, An ontology to semantically declare and describe functions, in: European Semantic Web Conference, 2016, pp. 46–49.
[14] A. C. Junior, C. Debruyne, R. Brennan, D. O'Sullivan, FunUL: a method to incorporate functions into uplift mapping languages, in: Proceedings of the 18th International Conference on Information Integration and Web-based Applications and Services, 2016, pp. 267–275.
[15] A. Iglesias-Molina, D. Van Assche, J. Arenas-Guerrero, B. De Meester, C. Debruyne, S. Jozashoori, P. Maria, F. Michel, D. Chaves-Fraga, A. Dimou, The RML Ontology: A Community-Driven Modular Redesign After a Decade of Experience in Mapping Heterogeneous Data to RDF, in: The Semantic Web – ISWC 2023: 22nd International Semantic Web Conference, Athens, Greece, November 6–10, 2023, Proceedings, Springer, 2023.
[16] M. Scrocca, A. Carenini, M. Grassi, M. Comerio, I. Celino, Not Everybody Speaks RDF: Knowledge Conversion between Different Data Representations, in: Proceedings of the

5th International Workshop on Knowledge Graph Construction, 2024.
[17] D. Van Assche, C. Debruyne, BURPing Through RML Test Cases, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[18] D. C. Herreros, D. Chaves-Fraga, M. Poveda-Villalón, R. Pernisch, L. Stork, O. Corcho, Propagating Ontology Changes to Declarative Mappings in Construction of Knowledge Graphs, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[19] E. de Vleeschauwer, P. Maria, B. De Meester, P. Colpaert, RML-view-to-CSV: A Proof-of-Concept Implementation for RML Logical Views, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2023.
[20] A. Randles, D. O'Sullivan, R2 [RML]-ChatGPT Framework, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[21] M. Hofer, J. Frey, E. Rahm, Towards Self-Configuring Knowledge Graph Construction Pipelines using LLMs - A Case Study with RML, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[22] C. Stadler, L. Bühmann, L.-P. Meyer, M. Martin, Scaling rml and sparql-based knowledge graph construction with apache spark., in: Proceedings of the 4th International Workshop on Knowledge Graph Construction (KGCW 2023), 2023.
[23] M. Freund, S. Schmid, R. Dorsch, A. Harth, FlexRML: A Flexible and Memory Efficient Knowledge Graph Materializer, in: The Semantic Web: 21st International Conference, ESWC 2024, Hersonissos, Crete, Greece, May 26–30, 2024, Proceedings, Part II, 2024.
[24] S. M. Oo, T. Verbeken, B. De Meester, RMLWeaver-JS: An algebraic mapping engine in the KGCW Challenge 2024, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[25] G. Haesendonck, W. Maroy, P. Heyvaert, R. Verborgh, A. Dimou, Parallel RDF Generation from Heterogeneous Big Data, in: Proceedings of the International Workshop on Semantic Big Data, 2019.
[26] E. de Vleeschauwer, B. De Meester, RMLStreamer supported by RML-view-to-CSV in the performance track of the KGCW Challenge 2024, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[27] M. Freund, S. Schmid, R. Dorsch, A. Harth, Performance Results of FlexRML in the KGCW Challenge 2024, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[28] D. Van Assche, J. Jankaj, B. De Meester, Backwards or Forwards? [R2]RML backwards compatibility in RMLMapper, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[29] C. Debruyne, D. Van Assche, The Conformance of an RML Processor Built from Scratch to Validate RML Specifications and Test Cases, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[30] E. Iglesias, M.-E. Vidal, Results for Knowledge Graph Creation Challenge 2024: SDM-RDFizer, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.
[31] C. Stadler, S. Bin, KGCW2024 Challenge Report: RDFProcessingToolkit, in: Proceedings of the 5th International Workshop on Knowledge Graph Construction, 2024.