Preface for the 6th Edition of the International Knowledge Graph Construction Workshop

David Chaves-Fraga[1], Ioannis Dasoulas[2,3,4], Christophe Debruyne[5], Anastasia Dimou[2,3,4], 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]*Montefiore Institute, University of Liège, Belgium*
[6]*Onlim, 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 has been considered an engineering task; however, more scientifically robust methods continue to emerge. These methods were widely questioned for their verbosity, low performance, or difficulty of use. At the same time, the variety and complexity of the data sources cause further issues with syntax and semantic interoperability.

Declarative methods (mapping languages) for describing rules to construct knowledge graphs, as well as approaches to execute those rules, continue to emerge. Nevertheless, constructing knowledge graphs remains a challenging task because several existing issues persist, and the barriers to their adoption are not sufficiently lowered to be easily and widely 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 develop groundbreaking solutions.

Addressing challenges related to knowledge graphs construction requires well-founded research, including the investigation of concepts and the 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 have been proposed [2, 3, 4]. Certain approaches followed the ETL-like paradigm, e.g., SDM-RDFizer [5], RocketRML [6], and FunMap [7], while 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].

In 2019, the W3C Knowledge Graph Construction Community Group[5] brought together researchers and practitioners with the overall goal of supporting its participants in developing better methods for Knowledge Graph construction. A significant milestone was reached in 2023, when a revised version of the RML specification [15] was published. Members of the community have also developed a performance benchmark [16], which have been used as part of the KGC challenge in 2023[6] and 2024[7].

Sixth International Workshop On Knowledge Graph Construction Co-located with the ESWC 2025, 1st June 2025, Portoroz, Slovenia
david.chaves@upm.es (D. Chaves-Fraga); ioannis.dasoulas@kuleuven.be (I. Dasoulas); c.debruyne@uliege.be (C. Debruyne); anastasia.dimou@kuleuven.be (A. Dimou); umutcan.serles@onlim.com (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
[5]https://www.w3.org/community/kg-construct/
[6]https://w3id.org/kg-construct/workshop/2023/challenge.html
[7]https://w3id.org/kg-construct//workshop/2024/challenge.html

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

The sixth edition of the knowledge graph construction workshop[8] was focused on the systematic assessment of various aspects of knowledge graph generation, including usability, usefulness, and coverage in terms of supported techniques, languages, and extensions, and the tradeoffs between various metrics and techniques in production settings. Thereof, this enabled the workshop to collect contributions from a wide range of topics such as the role of generative LLMs in (declarative) KG Generation, automation and planning of KG processes, and the role of human stakeholders in KG processes. It also included:

- *KGC Community Discussions.* This year, we experimented with engaging with the participants to distill and discuss two "outrageous" topics. With the help of an interactive platform to solicit topics and LLMs to categorize and combine them into two topics, the following two questions emerged: "If RML Is So Great, Why Does No One Want to Use It?" and "Is RML Falling Behind in a World of LLMs and Scalable Data Needs?"
  The conclusion of the first topic is that to gain traction, RML needs better usability, clearer value at smaller scales, and stronger ecosystem support. The second topic concluded that LLMs show promise in assisting with structure generation and document interpretation, but they currently fall short in producing accurate, deterministic RML mappings. This highlights the need for better hybrid workflows rather than full automation. This year's focus on users was thus timely.
- *The Third Knowledge Graph Construction Challenge.* Although the RML specification continues to evolve, no changes have necessitated a revision of the performance benchmark. We thus decided to focus this year's edition on the compliance with the 2025 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.

Ten papers were submitted. The reviews were open and public, and hosted at Open Review[9]. 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.

Six papers were accepted, and one was conditionally accepted. Six 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:

- *A Protocol for KG Construction Tasks Involving Users* [17]
- *Extending RML to Support Permissioned Data Sharing with Multiple Views* [18]
- *GRAPE: Guiding RML Authoring with a Projectional* [19]
- *On Dependencies in Knowledge Graph Construction* [20]
- *Mapping by Example: Towards an RML Mapping Reverse Engineering Pipeline* [21]
- *LLM-based Reranking and Validation of Knowledge Graph Completion* [22]
- *typhon-rml: Modularised Declarative Knowledge Graph Construction for Flexible Integrations and Performance Optimisation* [23]

During the workshop, the third edition of the Knowledge Graph Construction Challenge was organized, focusing on the conformance with the new RML modules[10]. The challenge was around conformance with the new RML modules, which encouraged developers of RML engines to support the specifications of the new RML modules by evaluating their engines against 337 test cases provided by the maintainers of each RML module. The core module, RML-Core (59 test cases), focuses on the core parts of RDF generation. RML-IO (73 test cases) focuses on input and output sources handling, while RML-IO-Registry (103 test cases) tests input source-specific configurations. Data transformations with FnO were also present through the RML-FNML (17 test cases) module. Newer modules,

[8]http://w3id.org/kg-construct/workshop/2025
[9]https://openreview.net/group?id=eswc-conferences.org/ESWC/2025/Workshop/KGCW
[10]https://w3id.org/rml/portal

e.g., RML-Star (18 test cases) for RDF-Star support, RML-CC (35 test cases) for generation of RDFS Collections & Containers, and RML-LV (32 test cases) for creating logical views on input data, provided new challenges for existing engines as they impact the RDF generation process. We had 4 participating engines: RMLMapper [2], SDM-RDFizer [5], RMLWeaver[11] and typhon-rml [23]

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

- *RMLMapper supported by RML-view-to-CSV in the KGCW Challenge 2025* [24]
- *Results for Knowledge Graph Creation Challenge 2025: SDM-RDFizer* [25]

Organizing Committee

- David Chaves-Fraga, Universidade de Santiago de Compostela
- Ioannis Dasoulas, KU Leuven
- Christophe Debruyne, University of Liège
- Anastasia Dimou, KU Leuven, Flanders Make, Leuven.AI
- Umutcan Serles, Onlim
- Dylan Van Assche, Ghent University – imec – IDLab

Program Committee

- Anelia Kurteva, Delft University of Technology
- Beatriz Esteves, Ghent University – imec – IDLab
- Ben De Meester, Ghent University – imec – IDLab
- Bram Steenwinckel, Ghent University – imec – IDLab
- 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 St George's, University of London
- 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
- Jakub Klímek, Charles University
- Juliette Opdenplatz, Universität Innsbruck
- Jürgen Umbrich, Vienna University of Economics and Business
- 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

[11]https://github.com/RMLio/rmlweaver-js

- 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, Sirma AI (Ontotext Corp)
- Ana Iglesias-Molina, BASF
- Diego Conde-Herreros, Universidad Politécnica de Madrid
- Romana Pernisch, Vrije Universiteit Amsterdam
- Eduard Kamburjan, IT University of Copenhagen
- Ignacio Domínguez Martínez-Casanueva, Telefónica
- Valentina Carriero, Cefriel
- Fajar Ekaputra, WU Vienna
- Laura Waltersdorfer, Vienna University of Technology

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. Crotti 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] D. Van Assche, D. Chaves-Fraga, A. Dimou, KROWN: A benchmark for RDF graph materialisation, in: G. Demartini, K. Hose, M. Acosta, M. Palmonari, G. Cheng, H. Skaf-Molli, N. Ferranti, D. Hernández, A. Hogan (Eds.), The Semantic Web - ISWC 2024 - 23rd International Semantic Web Conference, Baltimore, MD, USA, November 11-15, 2024, Proceedings, Part III, volume 15233 of *Lecture Notes in Computer Science*, Springer, 2024, pp. 20–39.
[17] A. Crotti Junior, C. Debruyne, A protocol for kg construction tasks involving users, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[18] E. de Vleeschauwer, G. Haesendonck, B. D. Meester, P. Colpaert, Extending rml to support permissioned data sharing with multiple views, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[19] J. Duchateau, C. Debruyne, Grape: Guiding rml authoring with a projectional, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[20] E. Kamburjan, R. Pernisch, O. Corcho, D. Chaves-Fraga, On dependencies in knowledge graph construction, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[21] M. Freund, R. Dorsch, S. Schmid, A. Hart, Mapping by example: Towards an rml mapping reverse engineering pipeline, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[22] W. Zhang, O. Serban, Llm-based reranking and validation of knowledge graph completion, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[23] M. Grassi, M. Scrocca, A. Carenini, I. Celino, typhon-rml: Modularised declarative knowledge graph construction for flexible integrations and performance optimisation, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[24] E. de Vleeschauwer, D. V. Assche, B. D. Meester, Rmlmapper supported by rml-view-to-csv in the kgcw challenge 2025, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.
[25] E. A. Iglesias, M.-E. Vidal, Results for knowledge graph creation challenge 2025: Sdm-rdfizer, in: Proceedings of the 6th International Workshop on Knowledge Graph Construction, 2025.