Preface for the 7th Edition of the International Knowledge Graph Construction Workshop

David Chaves-Fraga[1], Christophe Debruyne[2,†], Anastasia Dimou[3,4,5], Enrique Iglesias[6,7], Adrián Martinez-Balea[1] and Sitt Min Oo[8]

[1]*CiTIUS, Universidade de Santiago de Compostela, Santiago de Compostela, Spain*
[2]*Montefiore Institute, University of Liège, Belgium*
[3]*KU Leuven, Department of Computer Science, Sint-Katelijne-Waver, Belgium*
[4]*Flanders Make – DTAI-FET*
[5]*Leuven.AI – KU Leuven institute for AI, B-3000 Leuven, Belgium*
[6]*L3S Research Center, Hannover, Germany*
[7]*Leibniz University of Hannover, Hannover, Germany*
[8]*IDLab, Dept of Electronics and Information Systems, Ghent University – imec, Belgium*

More and more knowledge graphs (KG) 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 for automatically constructing KGs from existing Web objects exist, constructing KGs from heterogeneous data sources remains challenging. So far, KG construction has often been treated as an engineering task; however, more scientifically robust methods continue to emerge, including optimization, planning, query evaluation, mapping generation, and systematic evaluation. These methods have been widely "criticized for their verbosity, poor performance, and difficulty of use and reuse. At the same time, the variety and complexity of data sources pose further challenges to syntactic and semantic interoperability.

Declarative methods, such as mapping languages, specify rules for constructing KGs, while execution approaches materialize or virtualize the resulting graph. Nevertheless, defining these rules and constructing KGs is still not straightforward because several open questions remain: which system to use for which case, which system performs more accurately or efficiently, and when, and which system complies with which specification. These reasons, along with the large and active Knowledge Graph Construction W3C Community Group[4], indicate that there are still open questions requiring further investigation.

Addressing challenges related to KG construction requires well-founded research, including investigating concepts and developing tools and methods to evaluate them. R2RML was recommended by W3C in 2012, and since then, various extensions, alternatives, and implementations have been proposed [2, 3, 4]. Certain approaches followed the ETL-like paradigm, e.g., SDM-RDFizer [5, 6], and Morph-KGC [7], while others followed the query-answering paradigm, e.g., Morph-RDB [8] and Ontop [9]. Besides R2RML-based extensions, alternatives were proposed, e.g., SPARQL-Generate [10] and ShExML [11], as well as methods for performing data transformations during KG construction, e.g., FunMap [12] and FunUL [13].

In 2019, the W3C Knowledge Graph Construction Community Group brought together researchers and practitioners with the overall goal of supporting its participants in developing better methods for KG construction. A significant milestone was reached in 2023, when a revised version of the RML

Seventh International Workshop on Knowledge Graph Construction, co-located with ESWC 2026, May 10, 2026, Dubrovnik, Croatia
[†]C. Debruyne was partially supported by the Fonds de la Recherche Scientifique – FNRS under Grant n° MIS F.4016.24.
david.chaves@usc.es (D. Chaves-Fraga); c.debruyne@uliege.be (C. Debruyne); anastasia.dimou@kuleuven.be (A. Dimou); iglesias@l3s.de (E. Iglesias); adrianmartinez.balea@usc.es (A. Martinez-Balea); X.SittMinOo@UGent.be (S. M. Oo)

[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]https://www.w3.org/community/kg-construct/

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073
published 2026-08-24

specification [14] was published. Members of the community have also developed a performance benchmark [15], which was used as part of the KGC Challenge in 2023[5] and 2024[6]. This edition of the challenge has broadened the evaluation scope from performance to further dimensions of knowledge graph construction, including conformance with the latest RML modules and the methodology used to define, generate, and execute mappings.

The seventh edition of the Knowledge Graph Construction Workshop[7] focused on the systematic assessment of various aspects of knowledge graph generation, including usability, usefulness, coverage across supported techniques, languages, and extensions, and the trade-offs between various metrics and techniques in production settings. This enabled the workshop to collect contributions on a wide range of topics, including the role of generative LLMs in declarative KG generation, the automation and planning of KG processes, and the role of human stakeholders in KG processes. It also included *the Fourth Knowledge Graph Construction Challenge*, which covered three complementary dimensions of knowledge graph construction from heterogeneous data: conformance, performance, and mapping methodology.

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 KGs from academia and industry.

16 papers were submitted for peer review at this workshop. Of these, 13 papers were accepted for this volume: 7 as regular papers and 6 as short papers. The reviews were open and public, and hosted at OpenReview[8]. Each paper received at least three reviews from reviewers with different backgrounds and statuses.

The following workshop papers were accepted for publication and presented at the workshop:

- *Shifting Models Left: End-to-End Traceability as a Foundation for Knowledge Graphs as Data Products* [16]
- *Beyond Exit Code 1: Execution Report of RML Processors* [17]
- *From Federal Heterogeneity to Reproducible Analytics: A Provenance-Aware Knowledge Graph for Cross-Portal Comparability of German Open Government Data* [18]
- *RML Playground: Online Editing, Validation, and Execution for the RDF Mapping Language* [19]
- *Implementing the EHRI-KG: From Declarative Mapping Rules to a Three-Layered KG Construction Architecture* [20]
- *Towards Verifiable Declarative Mappings: A Vision Paper* [21]
- *Reducing Human Effort in Ontology-Based Data Integration: An LLM-Assisted Mapping Pipeline* [22]
- *Implementing the Semantic Units Framework using Named Graphs and R2RML in a Knowledge Graph for the Biodiversity Exploratories* [23]
- *Towards Standardised AI Documentation: A Knowledge Graph Approach for Model and Data Cards* [24]
- *NORKE: A Framework for Normative Rule-Centric Knowledge Graph Construction using Language Models* [25]
- *Let's Invert It! From RDF to Relational Data with R2RML* [26]
- *The HeFQUIN Approach to Extract and to Process Composite Values with RML and SPARQL* [27]

During the workshop, the fourth edition of the Knowledge Graph Construction Challenge was organized. Building on the insights gained from previous editions, this edition introduced three tracks that cover complementary dimensions of KG construction from heterogeneous data: conformance, performance, and mapping methodology. The challenge resources were made available through the Knowledge Graph Construction Challenge repository[9].

[5]https://w3id.org/kg-construct/workshop/2023/challenge.html
[6]https://w3id.org/kg-construct/workshop/2024/challenge.html
[7]https://w3id.org/kg-construct/workshop/2026
[8]https://openreview.net/group?id=eswc-conferences.org/ESWC/2026/Workshop/KGCW
[9]https://github.com/kg-construct/kgc-challenge

The first track focused on conformance with the new RDF Mapping Language (RML) modules developed by the W3C Knowledge Graph Construction Community Group. The goal was to evaluate the feasibility and correctness of the available test cases by applying them in existing implementations, and to foster the development of implementations for the new specifications. The track covered the following modules: RML-Core[10], RML-IO[11], RML-CC[12], RML-FNML[13], and RML-LV[14]. The official versions of the test cases used for the challenge were published on Zenodo for RML-Core[15], RML-IO[16], RML-CC[17], RML-FNML[18], and RML-LV[19]. RML-Star was not included in this year's challenge because RDF 1.2 evolved considerably and the specification needs to be adapted to the final recommendation.

The second track focused on performance. Besides execution time, the track considered computing resources such as CPU and memory usage, or combinations of these metrics. Participants were provided with virtual machines with uniform hardware resources to ensure a fair comparison of the mapping implementations. The performance challenge evaluated materialization capabilities under several scenarios. The data-oriented scenarios scaled the number of data records, the number of data properties, the percentage of duplicate values, and the percentage of empty values. The mapping-oriented scenarios scaled the numbers of subject maps, predicate-object maps, and named graphs, as well as the number and types of joins. This edition also introduced RML-CC collections and containers scenarios, including scaling the number of records with fixed-size lists, the number of values per list, the percentage of matching keys within lists, and the percentage of duplicated data within lists, as previously defined in [28].

The third track focused on mapping methodology. Although RML has become the de facto standard for constructing KGs from heterogeneous data sources, the design space for defining and executing mappings is far from closed. This track invited participants to propose novel solutions for KG generation, including approaches that build upon RML and its ecosystem, introduce extensions or optimizations, or depart from RML in favor of new mapping models, languages, automation techniques, or execution strategies. The focus was on innovation in how mappings are defined, generated, and executed, as well as on practical benefits such as reusability, maintainability, scalability, and expressiveness.

For the mapping methodology track, the organizers provided scenarios with input datasets in different formats, along with either an ontology that the output had to comply with or a target RDF graph to be generated from the input datasets. The first scenario was derived from the Public Procurement Data Space and provided public procurement notices extracted from the Tenders Electronic Daily platform in XML format. Participants had to generate an RDF KG compliant with a subset of the eProcurement Ontology, with the expected output graph provided in Turtle format. The second scenario was provided by meemoo[20] and required participants to transform JSON files into the expected RDF graphs, using the corresponding RDFS and SHACL documentation as reference.

The challenge received submissions from five engines in the conformance track: BURP-Error, Typhon-RML, CARML, MappingWeaver with RMLViewer, and SDM-RDFizer. Three engine variants participated in the performance track: SDM-RDFizer, CARML, and CARML-reactive. The conformance results showed that three of the five participating engines achieved full or near-full compliance with the five selected RML modules, while the performance track compared execution time, CPU usage, and memory consumption across the benchmark scenarios. The mapping methodology track did not yield a finalized submission; however, the Typhon-RML authors reported an exploratory agentic approach to automatically generating mapping rules in MTL, highlighting the need for clearer evaluation criteria

[10]https://w3id.org/rml/core/test-cases
[11]https://w3id.org/rml/io/test-cases
[12]https://w3id.org/rml/cc/test-cases
[13]https://w3id.org/rml/fnml/test-cases
[14]https://w3id.org/rml/lv/test-cases
[15]https://doi.org/10.5281/zenodo.19049860
[16]https://doi.org/10.5281/zenodo.19049774
[17]https://doi.org/10.5281/zenodo.19055451
[18]https://doi.org/10.5281/zenodo.19055650
[19]https://doi.org/10.5281/zenodo.19055778
[20]https://meemoo.be

for future editions, including functional correctness, readability, generalizability, and cost.

This edition also introduced the challenge participants to the possibility of submitting a single challenge report together. The report underwent peer review and was accepted for inclusion in the proceedings under the title *Technical Report on the 4th Edition of Knowledge Graph Construction Workshop Challenge* [29]. Furthermore, the challenge results of the first track were used as input for the RML Implementation Report[21], providing evidence of the implementation coverage of the RML specifications and their test cases across participating engines.

Organizing Committee

- David Chaves-Fraga, Universidade de Santiago de Compostela
- Christophe Debruyne, University of Liège
- Anastasia Dimou, KU Leuven – Flanders Make – Leuven.AI
- Enrique Iglesias, L3S Research Center – Leibniz University of Hannover
- Adrián Martínez-Balea, Universidade de Santiago de Compostela
- Sitt Min Oo, Ghent University – imec – IDLab

Program Committee

- Ademar Crotti Junior, metaphacts GmbH
- Ana Iglesias-Molina, Johnson and Johnson
- Anelia Kurteva, Delft University of Technology
- Beatriz Esteves, Ghent University – imec – IDLab
- Ben De Meester, Ghent University – imec – IDLab
- Davan Chiem Dao, University of Liège
- Davide Lanti, Free University of Bozen
- Dylan Van Assche, Independent Researcher
- Eduard Kamburjan, IT University of Copenhagen
- Els de Vleeschauwer, Ghent University
- Reza (Enayat) Rajabi, Cape Breton University
- Ernesto Jiménez-Ruiz, City St George's, University of London
- Fajar J. Ekaputra, WU Vienna
- Femke Ongenae, Ghent University – imec – IDLab
- 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 Duchateau, University of Liège
- Jakub Klímek, Charles University
- Lionel Tailhardat, Orange
- Manuel Lama, University of Santiago de Compostela
- Maria-Esther Vidal, Leibniz University of Hannover
- Mario Scrocca, Cefriel
- Markus Schröder, German Research Center for AI

[21] https://w3id.org/kg-construct/rml-implementation-report

- Michael Freund, Fraunhofer
- Oscar Corcho, Universidad Politécnica de Madrid
- Pieter Colpaert, Ghent University – imec – IDLab
- Samaneh Jozashoori, metaphacts GmbH
- Sergio José Rodríguez Méndez, Australian National University
- Sven Lieber, Royal Library Of Belgium
- Vladimir Alexiev, Graphwise
- Xuemin Duan, KU Leuven

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] E. Iglesias, M.-E. Vidal, D. Collarana, D. Chaves-Fraga, Empowering the SDM-RDFizer tool for scaling up to complex knowledge graph creation pipelines1, Semantic Web 16 (2025) SW–243580. URL: https://journals.sagepub.com/doi/abs/10.3233/SW-243580. doi:10.3233/SW-243580. arXiv:https://journals.sagepub.com/doi/pdf/10.3233/SW-243580.

[7] J. Arenas-Guerrero, D. Chaves-Fraga, J. Toledo, M. S. Pérez, O. Corcho, Morph-kgc: Scalable knowledge graph materialization with mapping partitions, Semantic Web 15 (2024) 1–20.

[8] 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.

[9] 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).

[10] M. Lefrançois, A. Zimmermann, N. Bakerally, A SPARQL Extension for Generating RDF from Heterogeneous Formats, in: The Semantic Web: 14th International Conference, 2017.

[11] 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.

[12] 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.

[13] 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.

[14] 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.
[15] D. Van Assche, D. Chaves-Fraga, A. Dimou, Krown: A benchmark for rdf graph materialisation, in: International Semantic Web Conference, Springer, 2024, pp. 20–39.
[16] L. Tailhardat, J. Balcerzack, A. Elmajed, P. Maria, G. Fomi Ngameni, Shifting models left: End-to-end traceability as a foundation for knowledge graphs as data products, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[17] J. Duchateau, D. Van Assche, C. Debruyne, Beyond exit code 1: Execution report of rml processors, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[18] F. Hahn, M. Martin, E. Engels, From federal heterogeneity to reproducible analytics: A provenance-aware knowledge graph for cross-portal comparability of german open government data, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[19] E. de Vleeschauwer, A. Vercruysse, B. De Meester, P. Colpaert, Rml playground: Online editing, validation, and execution for the rdf mapping language, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[20] H. García-González, M. Bryant, Implementing the ehri-kg: From declarative mapping rules to a three-layered kg construction architecture, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[21] E. Kamburjan, M. Pereira, Towards verifiable declarative mappings: A vision paper, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[22] S. Ashiddiiqi, F. J. Ekaputra, M. Z. Catur Candra, G. A. P. Saptawati Soekidjo, Reducing human effort in ontology-based data integration: An llm-assisted mapping pipeline, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[23] T. A. Mustafa, L. Vogt, T. Heger, B. Koenig-Ries, Implementing the semantic units framework using named graphs and r2rml in a knowledge graph for the biodiversity exploratories, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[24] A. Donald, E. Curry, I. Ullah, H. Chen, T. Iqbal, E. Muñoz, A. Galanopoulos, S. Saxena, Towards standardised ai documentation: A knowledge graph approach for model and data cards, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[25] E. de Souza, A. Gomes Regino, J. C. Dos Reis, Norke: A framework for normative rule-centric knowledge graph construction using language models, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[26] A. Massari, A. Dimou, Let's invert it! from rdf to relational data with r2rml, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[27] O. Hartig, The hefquin approach to extract and to process composite values with rml and sparql, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.
[28] C. Debruyne, S. Jaadari, D. Chaves-Fraga, Declarative generation of rdf collections and containers from heterogeneous data, 978-1-64368-616-5 (2025).
[29] S. M. Oo, E. Iglesias, J. Duchateau, M. Grassi, P. Maria, E. de Vleeschauwer, A. Martínez-Balea, M. Scrocca, M.-E. Vidal, Technical report on the 4th edition of knowledge graph construction workshop challenge, in: Proceedings of the 7th International Workshop on Knowledge Graph Construction, 2026.