

# Shifting Models Left: End-to-End Traceability as a Foundation for Knowledge Graphs as Data Products

Lionel Tailhardat<sup>1,\*</sup>, Joanna Balcerzak<sup>1</sup>, Arij Elmajed<sup>1,\*</sup>, Pano Maria<sup>2,\*</sup> and Gabriel Fomi-NGameni<sup>1</sup>

<sup>1</sup>Orange Research, France

<sup>2</sup>ModelDesk, The Netherlands

## Abstract

Various methods and tools already exist in the field of knowledge graph construction, which can be combined into pipelines for knowledge graph-based decision support systems. However, the semantic integration of multiple data sources into a knowledge graph is a craftsmanship task closely linked to the knowledge/data engineers' deep understanding of the information system they work in. This vision paper proposes to approach the problem from the opposite side of the pipeline, rather than assembling integration logic bottom-up, we start from the hypothesis of end-to-end traceability, the verification mechanism that makes "shifting semantics left" operational, and views the knowledge graph as a data product. Our proposals are grounded in the logical nature of Semantic Web technologies and the concept of formal proof of systems.

## Keywords

Semantic integration, Data lineage, Schema matching, Data mesh, RML, SPARQL, Formal proof

## 1. Introduction

Declarative approaches for Knowledge Graph Construction (KGC), such as the RDF Mapping Language (RML) [1] and related tools, have seen significant growth in recent years due to the enthusiastic involvement of scientific, technical, and industrial communities around having a common and logical framework to describe and implement data integration systems. Research in diverse related fields such as incremental data integration [2], ontology modifications propagation [3], the use of LLMs to produce KGC pipelines [4], and the cartography of Semantic Web-related tools [5] attests to this progress. This body of work forms a corpus of methods and tools that address fundamental needs for reliable semantic integration of heterogeneous data, thereby bringing the benefits of knowledge graphs to real-world knowledge bases and decision support systems [6, 7, 8].

Within this body of work, reflections on end-to-end and industry-grade KGC pipelines

---

*KGCW'26: 7th International Workshop on Knowledge Graph Construction, May 10, 2026, Dubrovnik, Croatia*

\*Corresponding author.

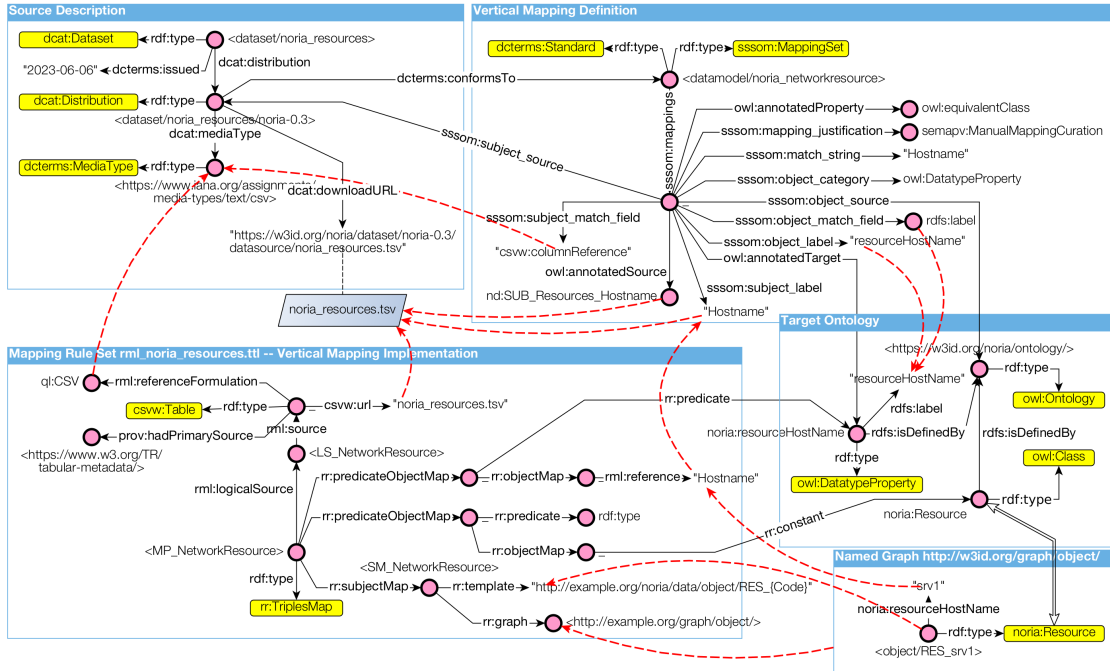
✉ lionel.tailhardat@orange.com (L. Tailhardat); joanna.balcerzak@orange.com (J. Balcerzak); arij.elmajed@orange.com (A. Elmajed); pano@modeldesk.io (P. Maria); gabriel.fomingameni@orange.com (G. Fomi-NGameni)

🆔 0000-0001-5887-899X (L. Tailhardat); 0009-0007-0396-1412 (J. Balcerzak); 0009-0007-2969-9141 (A. Elmajed); 0009-0000-2598-1894 (P. Maria); 0009-0004-7908-2463 (G. Fomi-NGameni)



© 2026 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)



**Figure 1:** Sketching end-to-end traceability in declarative data integration.

An attempt to connect the subgraphs involved in a KGC pipeline to understand how different frameworks (RML [1], SSSOM [18], DCAT [19], the NORIA-O [20] target ontology, etc.) interact, enable semantic mapping or data lineage calculations, and what connections might be missing (dashed lines) for end-to-end traceability and tackling the “semantic gap” issue. Five subgraphs are identified: data source description (top left), conceptual mapping (top right), mapping rules (bottom left), target ontology (right), and the resulting data graph (bottom right). Paths between subgraphs suggest alignment with the framework for understanding from Figure 2, such as `noria_resources.tsv` ← [Mapping Rule Set] → `srv1` for the “data lineage” case, and (Mapping Rule Set) ← [Source Description × Ontology] → (Vertical Mapping Definition) for the “Mapping templates” case. “Conceptual mapping” refers to high-level, implementation-agnostic correspondences between source schemas and target ontology concepts, as opposed to mapping rules (e.g., RML TriplesMap), which are their executable counterpart.

are approached through theoretical and methodological proposals, fundamentally relying on mathematical formalisms (e.g., algebraic [9, 10, 11], functional [12], category theory [13]) or practitioner interviews [14]. These works are instrumental in the design and deployment of KG-based systems and applications, particularly the mathematical approaches mentioned, which bring a model-based design and testing perspective to the field.

Nevertheless, when knowledge graphs are conceptualized as *data products* [15], through the combined lenses of semantic integration and data governance, it becomes evident that maintaining semantic continuity across the full KGC pipeline remains an unsolved challenge. Essential specification-level features<sup>1</sup> are still insufficiently defined, hindering a principled and reproducible approach to end-to-end KGC. For instance, in Figure 1, the lack of clear connection between data source characteristics and a set of RML rules disrupt provenance calculus [16, 17]. Throughout this paper, we use “data lineage” to denote the formal record of how each piece of information in the knowledge graph originates from and is transformed by upstream source data and mapping operations, distinct from data provenance, which typically refers to the origin

and custody of data rather than its full transformation trace.

Going a step further, in a *multi-domain KGC pipeline* context [21, 22], we observe that there is no standard way to express transformations made to the data at the knowledge graph level (a.k.a. patching queries [21]), nor how these relate to the upstream RML rules [23, 24]. To illustrate the implicit challenges of this, let us consider the case of building an IT Network Digital Twin (NDT) [25]: it involves integrating diverse network segments data (e.g., IP backbone, data centers, optical networks), each requiring different graph construction approaches (i.e., ways to map data elements and link them together) aligned with specific ontologies and domains. Despite a unified vocabulary (e.g. DevOpsInfra [26], NORIA-O [20], SEAS-PEP [27]), creating the NDT KG requires developing multiple domain-specific subgraphs and stitching them into a coherent, consistent and usable multi-domain knowledge graph for all stakeholders.

These observations converge with the Shift Left Data Manifesto [28], which highlights the challenge of maintaining semantic continuity from how data is structured (schemas) to its actual meaning in context (semantics). We refer to this as the *semantic gap* issue, and we argue that end-to-end traceability – from source data to user-level usage – is the operative mechanism for achieving this continuity and, therefore, a fundamental requirement for designing and using KG-based systems. Without it, semantic integration and knowledge sharing will remain a craftsmanship task closely tied to the knowledge/data engineer’s tacit understanding of the information system they work in.

In what follows, we further develop our vision by first sketching a framework for understanding and behavior-driven design of KGC pipelines (§2), and conclude by defining a high-level roadmap for future work (§3).

## 2. End-to-end Traceability as a Building Block

Based on the above discussion, it can be understood that five categories of objects need to be analyzed, both in their nature and in their relationships, to address the semantic gap issue and the multi-domain KGC challenges (Figure 2).

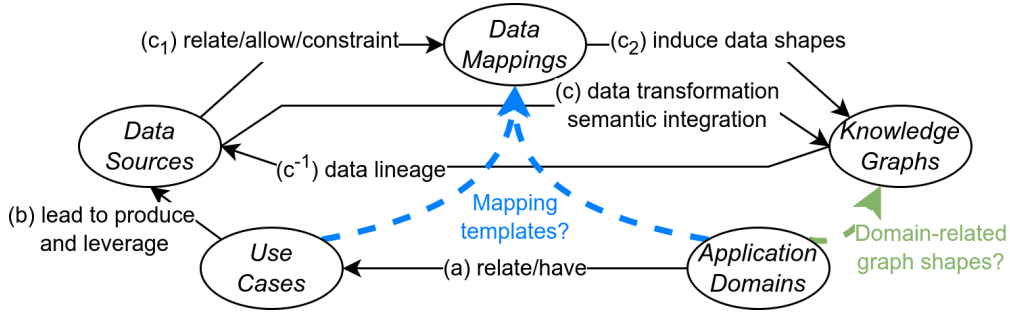
To further develop this intuition, we propose the following hypothesis as the foundation of our vision (**Hyp. 1 – End-to-End Traceability**): *end-to-end traceability should be a fundamental requirement for designing a KGC pipeline, if not considering a KG-based application.*

Practically, Hyp. 1 corresponds to demonstrating the following properties for any implementation of a multi-source, multi-process KGC pipeline: (**Prop. 1 – Lineage Completeness**) all source data contributing to an RDF triple are properly traced; (**Prop. 2 – Lineage Correctness**) the lineage graph [17, 29] does not contain spurious dependencies, that is, no source attribute is recorded as contributing to a triple unless a direct mapping path can be established between them; (**Prop. 3 – Lineage Minimality**) no unnecessary dependencies are added to the lineage graph.

In direct relation to the semantic gap issue, we further hypothesize (**Hyp. 2 – End-to-End Semantic**) that *end-to-end traceability should encompass a semantically analyzable description of source data (e.g., [30, 31, 32]) and a formal description of how end users leverage the KG*

---

<sup>1</sup>By which we mean properties of a KGC pipeline that can be expressed and verified independently of any specific implementation.



**Figure 2:** Objects involved in KGC using the category theory perspective.

With *Knowledge Graphs* as the target representation, we hypothesize that an *Application Domain* is related to (a edge) a set of *Use Cases*, which in turn involve using (b edge) a given set or type of *Data Sources*. These are shaped and combined through data transformation and semantic integration (c edge) in a convenient manner (e.g., running an RML engine and patching queries) to meet the requirements of the application domain and use cases. Looking backward at the relationships (i.e., treating  $c^{-1}$  as a modeled reciprocal of  $c_2 \circ c_1 = c$ , used here as an analytical device for formal verification rather than a claim of strict invertibility), we posit *Data Mappings* (e.g., RML TriplesMaps) to hold information on the use cases, representing some kind of functional dependency that enables revealing and studying *Mapping templates* and *Domain-related graph shapes* that could be generally applicable.

*data, including application domain description and implicit/explicit requirements (e.g., [33, 34]).* Practically, Hyp. 2 involves the KGC pipeline/KG-based application allowing the verification of the following two properties: (**Prop. 4 – Source Minimality**) identifying which source characteristics are mandatory for a given use case, and vice versa; (**Prop. 5 – Graph Shape**) determining the typical graph shapes [35] that can be expected from a set of sources with respect to a use case and/or an application domain. In particular, Prop. 4 directly operationalises the notion of data contracts [36]: a source attribute is mandatory if and only if it is declared as such in the contract governing the corresponding use case, making traceability not only a lineage concern but a governance one.<sup>2</sup>

Finally, in direct relation to the multi-domain KGC challenges and to support the scrutiny of Prop. 5, we consider that (**Hyp. 3 – Two Axes Mapping**) *the same concepts and properties of an ontology are involved in an orthogonal yet complementary manner during the KGC process, depending on whether it involves data transformation or semantic integration.* Practically, Hyp. 3 leads us to establish the following two definitions (which can be seen as lemmas for exploring the properties discussed above): **Vertical mapping** traverses abstraction levels within or across domains, connecting specific data-level representations to more general or more specialized ontological concepts, as in Figure 1 when RML rules lift tabular records into an OWL class hierarchy. **Horizontal mapping** operates across domains at equivalent levels of abstraction, establishing correspondences between conceptually similar entities in different ontological frameworks, analogous to a lateral alignment between parallel representations. This distinction impacts mapping reusability and transferability patterns, particularly relevant for understanding how declarative mapping rules can be effectively reused across different transformation contexts.<sup>3</sup>

<sup>2</sup>We remark that FAIR principles for KGs [37] provide an interesting additional enabling layer for Hyps. 1 & 2, as they offer the metadata infrastructure required for findable, accessible, interoperable, and reusable traceability artifacts across organizations and pipeline versions.

<sup>3</sup>The complementary roles of SSSOM and RML within this framework are noteworthy. SSSOM operates at the

### 3. Conclusion and Future Work

The end-to-end traceability hypothesis offers a principled basis for elevating KGC pipelines from craftsmanship-driven integration efforts to auditable data products, systems in which the relationship between source semantics, mapping logic, and graph output is formally verifiable. Systems satisfying Prop. 1 – 5 would address pressing operational needs: trustworthy multi-domain data lineage, reduced data collection footprints through source minimality, mapping rule inference grounded in explicit source and use-case descriptions rather than tacit expert knowledge.

Achieving these major objectives obviously requires experiments and intermediate implementations to further develop and test our vision. A first promising approach is to develop a *mathematical framework* that formalizes the usage context and explores data alignment via structure and semantic compatibility. For example, as suggested by [9], [11] and [38], datasets can be mathematically modeled and subjected to compatibility operations, enabling the algebraic analysis of data models. Closer to the KG realm, another option could be to address data alignment by comparing provenance information to then associate data [39, 40], whether it involves serial or parallel provenance (i.e., the chaining or parallelization of data integration traces). However, various levels of abstraction are to be explored – algebraic [9, 10, 11], functional [12], and category theory [13, 38]<sup>4</sup> – to identify the appropriate level of generalization.

A second promising approach is to challenge multi-domain KGC at design time using *formal proofs of systems* [41]. In an RML/SPARQL context, this involves validating data lineage properties by formalizing RML operations (initial generation of RDF triples), SPARQL operations (updates to the graph with patching queries), and the update rules for the lineage graph associated with these operations. Assuming that a mathematical framework has been chosen based on the scope of properties to be studied and the expected level of generalization, the challenge is not so much to prove the theoretical framework itself but to produce a faithful translation of the RML rules and SPARQL queries into this framework, and then to seek a proof of the properties for concrete instances of these translations.

A final promising approach is to explore *relationships with non-KG or legacy technologies*. For instance, the `datacontract-cli` tool from the ODCS project [36] offers extensive SQL tooling and a strong ecosystem, but focuses more on data access than semantics, unlike LinkML [18], which aligns with metadata registry concepts and includes model alignment tools. Translating structured representations from ontologies like DPROD [42] and LinkML-SSOM [18] into the `datacontract-cli` format could facilitate leveraging its connectors.

---

specification level, expressing semantic correspondences between concepts in a human-readable, ontology-agnostic format. In contrast, RML operates at the implementation level, encoding executable transformation logic. A specification should be substantially simpler than its implementation. In practice, an SSSOM mapping set for a given domain involves far fewer constructs than the corresponding RML TriplesMap, which must additionally encode source access patterns, value transformations, and named graph assignments. Hyp. 3 formalizes this distinction as the vertical/horizontal mapping axes.

<sup>4</sup>Since category theory enables modeling as much structure as the data has [13], an interesting avenue would be to explore how RML relates to category theory, considering RML as a model morphism [38] with an implicit product operator (i.e., composing data elements).

## Acknowledgments

We would like to thank Béatrice Rouxel, Yoan Chabot, Ivan Bedini, and Didier Bringer from Orange for their advice on the initial version of this project. We also thank Tobias Schweizer for early discussions on model crosswalks in relation to the MSCR project, and Vladimir Alexiev, Anelia Kurteva, and Ioannis Dasoulas (KGCW 2026 Program Committee members) for their appreciation and advice.

## References

- [1] A. Iglesias-Molina, D. Van Assche, J. Arenas-Guerrero, B. De Meester, C. Debruyne, S. Joza-shoori, 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*, volume 14266, Springer Nature Switzerland, Cham, 2023, pp. 152–175. doi:10.1007/978-3-031-47243-5\_9.
- [2] D. Van Assche, J. Andrés Rojas, B. De Meester, P. Colpaert, Incremental Knowledge Graph Construction from Heterogeneous Data Sources, *Semantic Web Journal* (2025). URL: <https://www.semantic-web-journal.net/content/incremental-knowledge-graph-construction-heterogeneous-data-sources-0>.
- [3] D. Conde-Herreros, L. Stork, R. Pernisch, M. Poveda-Villalón, O. Corcho, D. Chaves-Fraga, Propagating Ontology Changes to Declarative Mappings in Construction of Knowledge Graphs, in: *Proceedings of the 5<sup>th</sup> International Workshop on Knowledge Graph Construction Co-Located with 21<sup>th</sup> Extended Semantic Web Conference (ESWC)*, volume 3718, CEUR-WS.org, Hersonissos, Greece, 2024. URL: <https://ceur-ws.org/Vol-3718/paper1.pdf>.
- [4] M. Hofer, J. Frey, E. Rahm, Towards Self-Configuring Knowledge Graph Construction Pipelines Using LLMs - A Case Study with RML, in: *Proceedings of the 5<sup>th</sup> International Workshop on Knowledge Graph Construction Co-Located with 21<sup>th</sup> Extended Semantic Web Conference (ESWC)*, CEUR-WS.org, Hersonissos, Crete, Greece, 2024. URL: <https://ceur-ws.org/Vol-3718/paper6.pdf>.
- [5] A. Reiz, F. J. Ekaputra, N. Mihindikulasooriya, Semantic Tool Hub: Towards a Sustainable Community-Driven Documentation of Semantic Web Tools, in: *The Semantic Web: ESWC 2024 Satellite Events*, volume 15344, Springer Nature Switzerland, Cham, 2025, pp. 311–315. doi:10.1007/978-3-031-78952-6\_48.
- [6] M. Muller, European Parliament Open Data Portal : A SHACL-Powered Knowledge Graph, 2025. URL: <https://blog.sparna.fr/2025/04/09/european-parliament-open-data-portal-a-shacl-powered-knowledge-graph/>.
- [7] P. Lisena, D. Schwabe, M. Van Erp, R. Troncy, W. Tullett, I. Leemans, L. Marx, S. C. Ehrich, Capturing the Semantics of Smell: The Odeuropa Data Model for Olfactory Heritage Information, in: *The Semantic Web*, volume 13261, Springer International Publishing, Cham, 2022, pp. 387–405. doi:10.1007/978-3-031-06981-9\_23.
- [8] L. Tailhardat, Y. Chabot, A. Py, P. Guillemette, NORIA UI: Efficient Incident Management on Large-Scale ICT Systems Represented as Knowledge Graphs, in: *19<sup>th</sup> International*

- Conference on Availability, Reliability and Security (ARES), ARES '24, Association for Computing Machinery, New York, NY, USA, 2024. doi:10.1145/3664476.3670438.
- [9] S. M. Oo, B. De Meester, R. Taelman, P. Colpaert, Algebraic Mapping Operators for Knowledge Graph Generation (2025). URL: <https://www.semantic-web-journal.net/content/algebraic-mapping-operators-knowledge-graph-generation-0>.
- [10] T. J. Green, G. Karvounarakis, V. Tannen, Provenance Semirings, in: Proceedings of the Twenty-Sixth ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems (PODS), Association for Computing Machinery, New York, NY, USA, 2007, pp. 31–40. doi:10.1145/1265530.1265535.
- [11] M. Lenzerini, Data Integration: A Theoretical Perspective, in: Proceedings of the Twenty-First ACM SIGMOD-SIGACT-SIGART Symposium on Principles of Database Systems, ACM, Madison Wisconsin, 2002, pp. 233–246. doi:10.1145/543613.543644.
- [12] S. Geisler, C. Cappiello, I. Celino, D. Chaves-Fraga, A. Dimou, A. Iglesias-Molina, M. Lenzerini, A. Rula, From Genesis to Maturity: Managing Knowledge Graph Ecosystems Through Life Cycles 18 (2025). URL: <https://www.vldb.org/pvldb/vol18/p1390-geisler.pdf>.
- [13] V. Uotila, J. Lu, A Formal Category Theoretical Framework for Multi-Model Data Transformations, in: Heterogeneous Data Management, Polystores, and Analytics for Healthcare, volume 12921, Springer International Publishing, Cham, 2021, pp. 14–28. doi:10.1007/978-3-030-93663-1\_2.
- [14] R. Pernisch, D. Chaves-Fraga, L. Stork, D. Conde-Herreros, M. Poveda-Villalón, LOT4KG: A Joint Methodology for the Ontology and Knowledge Graph Lifecycle (2025).
- [15] J. Majchrzak, S. Balnojan, M. Siwiak, M. Sierackiewicz, Data Mesh in Action, Manning Publications Co, Shelter Island, NY, 2023.
- [16] M. Herschel, R. Diestelkämper, H. Ben Lahmar, A Survey on Provenance: What for? What Form? What From? 26 (2017) 881–906. doi:10.1007/s00778-017-0486-1.
- [17] T.-D. Trinh, P. R. Aryan, B.-L. Do, F. J. Ekaputra, E. Kiesling, A. Rauber, P. Wetz, A. M. Tjoa, Linked Data Processing Provenance: Towards Transparent and Reusable Linked Data Integration, in: Proceedings of the International Conference on Web Intelligence, ACM, Leipzig Germany, 2017, pp. 88–96. doi:10.1145/3106426.3106495.
- [18] N. Matentzoglou, et al., A Simple Standard for Sharing Ontological Mappings (SSSOM) (2022). doi:10.1093/database/baac035.
- [19] F. Maali, J. Erickson, Data Catalog Vocabulary (DCAT), W3C Recommendation SPSPD-vocab-dcat-20200204, W3C, 2020. URL: <https://www.w3.org/TR/vocab-dcat/>.
- [20] L. Tailhardat, Y. Chabot, R. Troncy, NORIA-O: An Ontology for Anomaly Detection and Incident Management in ICT Systems, in: The Semantic Web, Springer Nature Switzerland, Cham, 2024, pp. 21–39. doi:10.1007/978-3-031-60635-9\_2.
- [21] L. Tailhardat, Y. Chabot, R. Troncy, Designing NORIA: a Knowledge Graph-based Platform for Anomaly Detection and Incident Management in ICT Systems, in: 4<sup>th</sup> International Workshop on Knowledge Graph Construction (KGCW), volume 3471, CEUR-WS.org, 2023. URL: <https://ceur-ws.org/Vol-3471/paper3.pdf>.
- [22] D. Greco, F. Osborne, S. Puseddu, D. Reforgiato Recupero, Modelling Big Data Platforms as Knowledge Graphs: The Data Platform Shaper, Journal of Big Data 12 (2025). doi:10.1186/s40537-025-01094-w.
- [23] 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 5<sup>th</sup> International Workshop on Knowledge Graph Construction Co-Located with 21<sup>th</sup> Extended Semantic Web Conference (ESWC 2024), volume 3718, CEUR-WS.org, Hersonissos, Greece, 2024. URL: <https://ceur-ws.org/Vol-3718/paper2.pdf>.
- [24] E. Kamburjan, R. Pernisch, O. Corcho, D. Chaves-Fraga, On Dependencies in Knowledge Graph Construction, in: Proceedings of the 6<sup>th</sup> International Workshop on Knowledge Graph Construction co-located with 22<sup>nd</sup> Extended Semantic Web Conference (ESWC), volume 3999, CEUR-WS.org, 2025. URL: <https://ceur-ws.org/Vol-3999/paper4.pdf>.
- [25] D. Allison, P. Smith, K. McLaughlin, Digital Twin-Enhanced Incident Response for Cyber-Physical Systems, in: Proceedings of the 18<sup>th</sup> International Conference on Availability, Reliability and Security (ARES), Association for Computing Machinery, 2023. doi:10.1145/3600160.3600195.
- [26] O. Corcho, D. Chaves-Fraga, J. Toledo, J. Arenas-Guerrero, C. Badenes-Olmedo, M. Wang, H. Peng, N. Burrett, J. Mora, P. Zhang, A High-Level Ontology Network for ICT Infrastructures, in: 20<sup>th</sup> International Semantic Web Conference (ISWC), 2021. doi:10.1007/978-3-030-88361-4\_26.
- [27] M. Lefrançois, J. Kalaoja, T. Ghariani, A. Zimmermann, SEAS Knowledge Model, Deliverable 2.2, ITEA2 12004 Smart Energy Aware Systems, 2016. URL: <https://hal.science/hal-01885354>.
- [28] C. Sanderson, The Shift Left Data Manifesto, <https://www.gable.ai/blog/shift-left-data-manifesto>, 2025.
- [29] A. Dimou, T. D. Nies, R. Verborgh, E. Mannens, R. Van de Walle, Automated Metadata Generation for Linked Data Generation and Publishing Workflows, in: WWW2016 Workshop: Linked Data on the Web (LDOW2016), Montreal Quebec Canada, 2016. URL: [http://events.linkedata.org/ldow2016/papers/LDOW2016\\_paper\\_04.pdf](http://events.linkedata.org/ldow2016/papers/LDOW2016_paper_04.pdf).
- [30] M. Fiorelli, A. Stellato, T. Lorenzetti, A. Turbati, W. van Gemert, D. Dechandon, A. Gerencsér, E. Francesconi, MDR: An Ontology Vocabulary and Registry Service for Dataset Catalogs, in: Advances and Trends in Artificial Intelligence. Theory and Applications, Springer Nature Singapore, Singapore, 2026, pp. 354–366. doi:10.1007/978-981-96-8889-0\_31.
- [31] V. Mouche, M. Munier, N. Laamech, MMO: Une Ontologie Pour Les Métadonnées, *Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances*, RNTI-E-42 (2026) 371–372.
- [32] R. Pollock, J. Tennison, G. Kellogg, I. Herman, Metadata Vocabulary for Tabular Data, W3C Recommendation REC-tabular-metadata-20151217, W3C, 2015. URL: <https://www.w3.org/TR/tabular-metadata/>.
- [33] M. Ortega, H. Imhah, V. Khatchatrian, N. Mellouli, C. Rodrigues, N. Travers, Text2stories : Extraction et génération de user stories à partir de documentation fonctionnelle et non-fonctionnelle, *Revue des Nouvelles Technologies de l'Information Extraction et Gestion des Connaissances*, RNTI-E-42 (2026) 513–520.
- [34] E. Dubois, P. Du Bois, A. Rifaut, Elaborating, Structuring and Expressing Formal Requirements of Composite Systems, in: Advanced Information Systems Engineering, Springer, Berlin, Heidelberg, 1992, pp. 327–347. doi:10.1007/BFb0035140.
- [35] H. J. Pandit, D. O'Sullivan, D. Lewis, Using Ontology Design Patterns To Define SHACL

- Shapes, in: Proceedings of the 9<sup>th</sup> Workshop on Ontology Design and Patterns (WOP) Co-Located with 17<sup>th</sup> International Semantic Web Conference (ISWC), volume 2195, CEUR-WS.org, 2018. URL: [https://ceur-ws.org/Vol-2195/research\\_paper\\_3.pdf](https://ceur-ws.org/Vol-2195/research_paper_3.pdf).
- [36] Bitol, Open Data Contract Standard (ODCS), LF AI & Data Foundation, 2025. URL: <https://bitol-io.github.io/open-data-contract-standard>.
- [37] A. Serra, FAIR Principles for Knowledge Graphs in SSbD, 2025. doi:10.5281/zenodo.17305647, 13<sup>th</sup> Venice Training School, Ca' Foscari University of Venice, Italy.
- [38] J. Shinavier, R. Wisnesky, J. G. Meyers, Algebraic property graphs, 2022. doi:10.48550/arXiv.1909.04881.
- [39] F. Cotton, G. Duffes, F. Rizzolo, Using PROV-O to Represent Lineage in Statistical Processes: A Record Linkage Example, in: Joint Proceedings of the International Workshops on Sensors and Actuators on the Web, and Semantic Statistics Co-Located with 18<sup>th</sup> International Semantic Web Conference (ISWC), volume 2549, CEUR-WS.org, 2019. URL: <https://ceur-ws.org/Vol-2549/article-08.pdf>.
- [40] I. M. Yazici, E. Karabulut, M. S. Aktas, A Data Provenance Visualization Approach, in: 14<sup>th</sup> International Conference on Semantics, Knowledge and Grids (SKG), 2018. doi:10.1109/SKG.2018.00019.
- [41] I. Sergey, Programs and Proofs: Mechanizing Mathematics with Dependent Types, 2014. doi:10.5281/zenodo.4996238, lecture notes with exercises.
- [42] A. Gioia, J. Geluk, B. W. Smith, S. Fisher, O. Bage, B. Clinch, C. Mattocks, C. Tubbax, C. Ivie, Data Product Ontology (DPROD), 2025. URL: <https://ekgf.github.io/dprod/>.