Mapping Workbench: A collaborative platform for mapping complex XML data to RDF

Eugeniu Costetchi[1], Jana Ahmad[2], Csongor I. Nyulas[3], Rashif Rahman[4] and Dumitru Prijilevschi[5]

[1]Meaningfy SARL, 61 Route de Fischbach, Lintgen, L-7447, Luxembourg

Abstract

Mapping Workbench (MWB) facilitates the transformation of XML data into RDF using the RML mapping language, based on a sound test-driven methodology. This paper presents MWB as a comprehensive solution for Semantic Engineers and Data Modellers, offering efficient data mapping, management and validation against ontologies and data shapes. We discuss MWB's key features, the mapping process, and its potential impact on the fields of knowledge graph construction, semantic data interoperability and data integration in general.

Keywords

XML, RML, RDF, OWL, Data Mapping, Data Transformation, Knowledge Graph Construction

1. Introduction

Semantic data integration and transformation are crucial tasks in various domains involving knowledge graph (KG) construction, including bio-informatics, healthcare, finance, government open data and beyond. Mapping XML data against ontologies (and data shapes) is a common requirement in such domains. It enables data harmonisation, interoperability, knowledge sharing, and semantic data enrichment. While much attention was given to development of declarative mapping languages like RML[1] (an extension of[1] R2RML) and transformation engines to execute the transformation rules, little work has been done on tools that facilitate mapping rules development, testing, and management efficiently and accurately.

Mapping Workbench (MWB) [2] addresses this need by providing a user-friendly graphical user interface (GUI) and powerful functionalities for developing and testing RML mapping rules, managing complexity of large data structures, and collaborative editing and validating with domain experts, in the manner of an Integrated Development Environment (IDE). By adopting a test-driven approach for knowledge engineering and incorporating the notion of Conceptual Mapping (CM), MWB allows for evaluating the correctness of both the mapping rules and the transformed data against OWL/RDFS ontologies, SHACL data shapes and semi-automatically

SEMANTiCS 2024: 20th International Conference on Semantic Systems, September 17-19, 2024, , Amsterdam

*Corresponding author.

†These authors contributed equally.

eugen@meaningfy.ws (E. Costetchi); jana.ahmad@meaningfy.ws (J. Ahmad); csongor.nyulas@meaningfy.ws (C. I. Nyulas); rashif.rahman@meaningfy.ws (R. Rahman); dumitru.prijilevschi@meaningfy.com (D. Prijilevschi)

[1] https://rml.io/docs/rml/rmlvsr2rml/

[2] Mapping Workbench Home Page, Mapping Workbench Demo Application

generated validation rules derived from the CM.

2. Related Work

MWB is positioned as an IDE for RML mapping and therefore also finds itself in the space for RML GUI (web) applications (apps). **RMLEditor** [2] provides a simple, graph-based visual editing web app for domain experts to model knowledge from different data sources, using RML under the hood. However, managing complex data and mappings can become challenging with its basic GUI and limited RML features. **Karma** [3] is a more advanced ”information integration tool”, with comprehensive functionalities for loading data from multiple sources and automatic model alignment. Its tabular approach can complicate the creation of mappings and make interlinking between tables unnecessarily complex. **Ontopic**[3] Studio is a ”low-code” front-end to the **Ontop** [4] Virtual Knowledge Graph (VKG) system, which exposes (primarily relational) data dynamically as an active RDF graph ready for querying, without materializing the transformation. It translates SPARQL queries into SQL using R2RML mappings, and support for tree-structured source input formats (XML, JSON) is absent. **RMLx** [5] is yet another GUI for RML, but it facilitates the mapping with form-based input, which does little to alleviate mental effort (something GUIs should aim to do). Although MWB does not prioritize visual editing, it aims to incorporate automation and complexity management features that will ease the user’s cognitive load.

What is Mapping Workbench?

MWB is designed to simplify the complex task of converting XML data to RDF, involving model mapping between XML schemes [6] and RDFS/OWL ontologies [7]. It aims to improve efficiency and accuracy in large-scale data mapping projects by bringing together all necessary resources in one place. This integration streamlines the entire process of mapping development and management, starting from the initial planning stages to the final deployment or dissemination. By involving business stakeholders early on through writing Conceptual Mapping (CM) rules, MWB ensures that domain interpretations and practical needs align smoothly with eventual technical implementation, which helps to minimize unnecessary revisions and costs.

MWB also effectively handles the challenges posed by evolving XML schemes across revisions, ensuring that mapping rules are created with high precision across data versions through rigorous validation processes. This collaborative platform encourages teamwork between domain experts and Semantic Engineers, supported by role-based access controls that maintain strict data security and integrity standards. Its structured four-stage mapping approach includes (i) Conceptual Mapping using “ontology fragments”,[4] (ii) Technical Mapping using RML rules [1], (iii) validation using SHACL shapes [9], SPARQL assertions [10], and XPath queries [11], and (iv) the export of mapping packages or *suites* for seamless deployment into data transformation workflows. Users of MWB benefit from its sound methodology [12], which distinguishes between Conceptual and Technical Mappings, automated quality checks, and a user-friendly

[3] https://ontop-vkg.org/

[4] A custom dialect of SPARQL Path patterns [8] that includes the classes of intermediary nodes.

interface conducive to agile workflows. These features collectively enhance efficiency, reduce mapping time and costs, and increase satisfaction among stakeholders.

3. Features and Workflow

Figure 1 shows the MWB workflow, where each step has a set of functionalities that facilitate the mapping process.

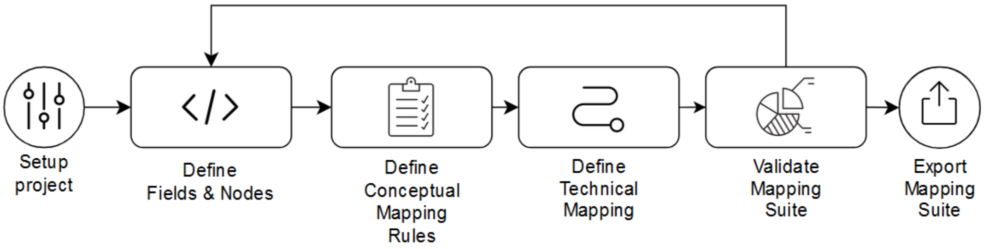

Figure 1: MWB Workflow

The workflow begins with the **Project Setup**, which includes adding test data, ontologies and other resources. There is a convenient interface for defining XML elements based on their XPaths. For ontologies, MWB offers an automatic detection mechanism of ontology terms. The next step is **Defining Conceptual Mappings**. MWB provides a user-friendly interface for determining correspondences between the elements from input data and the target terms from an ontology, allowing users to select models and view already created rules. In the **Technical Mapping Definition** step, the user imports or writes RML rules implementing what is designed and specified by the Conceptual Mapping Rules. The user is also able to transform (via RMLMapper [13]) one or more test files to observe output in short validation cycles. The **Mapping Suite Validation** in MWB involves a set of automatic tools that generates the mapping results, analyses them, and displays a set of reports. These reports include views with statistics and messages to support experts in making decisions about the correctness of the mappings. The reports are divided into:

- SHACL Report - shows the validity of data according to SHACL constraints.
- SPARQL Report - helps to understand the correctness of defined CM Rules.
- XPath Report - a detailed view that shows the coverage of the XPaths from test data.

At this point, MWB facilitates an iterative process of making changes, transforming them, and analysing the results. Then, the mapping suite can be **Exported** in an archived (ZIP) format.

4. Innovative Aspects

The platform presents several groundbreaking features not commonly found in other mapping development tools. Notably, the integration of ontologies, sample test data, mapping rules, and validation mechanisms into a single, cohesive platform marks a significant innovation in this field. A robust methodology has been developed for the creation and management of mapping rules throughout the mapping lifecycle. This structured approach enhances the accuracy and effectiveness of the mapping process. Moreover, the platform simplifies the mapping configuration and execution processes, enabling users to define mappings, establish rules, and transform data with minimal technical expertise required. This user-friendly approach lowers the barrier to entry for knowledge engineers and domain experts engaging with complex data transformation tasks. A particularly novel aspect is the bifurcation of mapping development into two distinct layers: Conceptual Mapping and Technical Mapping. This dual-layer structure maintains domain experts involved through an intuitive user interface, which facilitates their assessment of mapping rules for domain-specific soundness. Furthermore, CMs provide a basis for generating unit tests, thereby supporting the rigorous testing of technical rule implementations. This innovative separation not only enhances the manageability of mappings but also ensures their relevance and accuracy by involving domain knowledge at every step.

5. Benefits and Future Directions

MWB guarantees XML-RDF data mapping quality by measuring mapping validity, accuracy, and coverage. It significantly speeds up the mapping process and reduces maintenance costs. The platform identifies conceptual mapping issues early by involving domain experts, handling the complexity of large schemes, contextual mappings, and evolving schema versions. MWB provides a collaborative mapping environment for semantic engineers and domain experts, and prepares data for advanced KG, ML and AI applications, unlocking new use cases.

MWB aims to evolve into a dynamic Software as a Service (SaaS) platform, serving comprehensively domain experts and semantic engineers. Planned enhancements include advanced RML editing capabilities, generation of RML rules form the CM rules, further automation including GenAI-assisted mapping, generalized support for tree-structured data by way of expanded support for mapping JSON schemes, and continuous improvements in user interface and overall user experience. These advancements aim to make data integration efforts easier, strengthening MWB's position as a top solution that addresses technical data mapping challenges and promotes transparency, and adherence to evolving data interoperability standards.

6. Conclusion

Mapping Workbench stands at the forefront of semantic data integration, offering a framework to map complex XML data to RDF with unparalleled efficiency and accuracy. By centralising resources and fostering collaboration between stakeholders, MWB bridges the gap between conceptual understanding and technical implementation, significantly optimizing the mapping development life-cycle. Its structured approach ensures high-quality mapping rules, validated through sophisticated mechanisms like SHACL, SPARQL, and XPath, leading up to a seamless deployment in data transformation pipelines.

Embrace the future of semantic mapping with MWB — where efficiency, accuracy, and innovation converge to redefine data integration.

References

[1] A. Dimou, M. Vander Sande, P. Colpaert, R. Verborgh, E. Mannens, R. Van de Walle, RML: a generic language for integrated RDF mappings of heterogeneous data, in: C. Bizer, T. Heath, S. Auer, T. Berners-Lee (Eds.), Proceedings of the 7th Workshop on Linked Data on the Web, volume 1184 of *CEUR Workshop Proceedings*, 2014. URL: http://ceur-ws.org/Vol-1184/ldow2014_paper_01.pdf.

[2] P. Heyvaert, A. Dimou, A.-L. Herregodts, R. Verborgh, D. Schuurman, E. Mannens, R. Van de Walle, Rmleditor: A graph-based mapping editor for linked data mappings, in: H. Sack, E. Blomqvist, M. d'Aquin, C. Ghidini, S. P. Ponzetto, C. Lange (Eds.), The Semantic Web. Latest Advances and New Domains, Springer International Publishing, Cham, 2016, pp. 709–723.

[3] C. Knoblock, P. Szekely, J. L. Ambite, A. Goel, S. Gupta, K. Lerman, M. Muslea, M. Taheriyan, P. Mallick, Semi-automatically mapping structured sources into the semantic web, volume 7295, 2012. doi:10.1007/978-3-642-30284-8_32.

[4] M. Rodriguez-Muro, J. Hardi, D. Calvanese, Quest: Efficient sparql-to-sql for rdf and owl, volume 914 of *CEUR Workshop Proceedings*, RWTH, Aachen, 2012, pp. 53 – 56.

[5] P. Aryan, F. Ekaputra, K. Kurniawan, E. Kiesling, A. M. Tjoa, Rmlx : Mapping interface for integrating open data with linked data exploration environment, 2017. doi:10.31227/osf.io/qhdc9.

[6] T. Bray, J. Paoli, C. M. Sperberg-McQueen, E. Maler, F. Yergeau, Extensible markup language (xml) 1.0 (fifth edition), W3C Recommendation, 2008. Available at http://www.w3.org/TR/REC-xml/.

[7] S. Bechhofer, F. van Harmelen, J. Hendler, I. Horrocks, D. McGuinness, P. Patel-Schneijder, L. A. Stein, OWL Web Ontology Language Reference, Recommendation, World Wide Web Consortium (W3C), 2004. See http://www.w3.org/TR/owl-ref/.

[8] SPARQL 1.1 Query Language, Technical Report, W3C, 2013. URL: http://www.w3.org/TR/sparql11-query.

[9] Shapes constraint language (SHACL), Technical Report, W3C, 2017. URL: https://www.w3.org/TR/shacl/.

[10] E. Prud'hommeaux, A. Seaborne, SPARQL Query Language for RDF, W3C Recommendation, 2008. URL: http://www.w3.org/TR/rdf-sparql-query/.

[11] J. Clark, S. J. DeRose, Xml path language (xpath) version 1.0, World Wide Web Consortium, Recommendation REC-xpath-19991116, 1999.

[12] E. Costetchi, A. Vassiliades, C. I. Nyulas, Towards a mapping framework for the tenders electronic daily standard forms., in: KGCW@ ESWC, 2023.

[13] A. Dimou, T. De Nies, R. Verborgh, E. Mannens, P. Mechant, R. Van de Walle, Automated metadata generation for linked data generation and publishing workflows, in: LDOW2016, CEUR-WS. org, 2016, pp. 1–10.