OpenAPI to RDF: Automated Generation of RDF Vocabularies and SHACL Validation Shapes from OpenAPI Schema

Jean Martins [1], **Leonid Mokrushin** [1], **and Marin Orlic** [1]

1 Ericsson Research

Summary

openapi-to-rdf is a Python package that automatically converts data schemas from OpenAPI specifications into semantic web formats. OpenAPI is a widely-used standard for describing web APIs and data structures, commonly used in telecommunications, cloud computing, and web services. While OpenAPI specifications include API endpoints, parameters, and responses, this tool specifically focuses on the schema definitions that describe the structure and constraints of data models.

The package transforms these schema definitions from OpenAPI YAML files into two complementary outputs: RDF (Resource Description Framework) vocabularies that define concepts and their relationships, and SHACL (Shapes Constraint Language) validation shapes that enable automated data validation. This conversion bridges the gap between syntactic schema definitions and semantic web technologies, enabling automated reasoning, validation, and integration across different systems.

The converter handles all major OpenAPI schema features including nested objects, arrays, enumerations, and complex logical relationships (oneOf, anyOf, allOf). It resolves cross-references between multiple specification files and generates globally unique identifiers for all schema components. The package provides both command-line tools and a Python API, with comprehensive quality assurance reporting.

Testing on 34 telecommunications specifications shows 100% conversion success, processing over 2,000 data structures and 5,000 properties with complete preservation of constraints and documentation. The tool has enabled automated validation systems and semantic integration projects in the telecommunications industry.

Statement of Need

Modern software systems increasingly rely on APIs (Application Programming Interfaces) to communicate and share data. OpenAPI has emerged as the standard for documenting these APIs, with thousands of specifications available across industries. While OpenAPI specifications describe complete APIs including endpoints and operations, the schema definitions within these specifications – which define data structures, validation rules, and relationships—exist in isolation without explicit semantic relationships to other schemas, making integration and automated processing difficult.

The telecommunications industry exemplifies this challenge. The 3rd Generation Partnership Project (3GPP) publishes hundreds of OpenAPI specifications defining network functions and data models (3GPP, 2023). These specifications contain rich schema definitions describing network elements, policies, and service data structures. Meanwhile, industry frameworks like

TM Forum's Intent-Based Automation require semantic representations using RDF and SHACL for automated network management (TM Forum, 2024a, 2024b). Currently, bridging between these syntactic schema definitions and semantic frameworks requires manual, error-prone translation.

Existing tools for OpenAPI processing focus on code generation, documentation, or testing, but none systematically convert schema definitions to semantic web formats. Academic research has explored individual aspects of API-to-RDF conversion but lacks comprehensive tooling for real-world schema specifications with complex dependencies and domain-specific requirements.

The `openapi-to-rdf` package addresses this gap by providing automated conversion from OpenAPI schema definitions to W3C-standard semantic formats (W3C, 2014, 2017). This enables:

- **Intent-based automation**: Enable semantic representation of telecom network models for TM Forum Intent-Based Automation frameworks
- **Domain knowledge integration**: Convert 3GPP information models from OpenAPI schemas to RDF vocabularies to extend TM Forum Intent Ontology (TIO)
- **Automated SHACL validation**: Generate SHACL validation shapes that enable automated verification of intent compliance against domain specific 3GPP models

Therefore, the tool has particular relevance for telecommunications, where it enables intent-based network management systems, but applies broadly to any domain using OpenAPI specifications including cloud platforms, IoT systems, and web services.

References

3GPP. (2023). *3rd generation partnership project (3GPP) technical specifications*. Technical Specification Group Services and System Aspects (SA). https://www.3gpp.org/specifications

TM Forum. (2024a). *Intent common model – vocabulary reference (TR290v) v3.2.0*. TM Forum Technical Report. https://www.tmforum.org/resources/introductory-guide/tr290v-intent-common-model-vocabulary-reference-v3-2-0/

TM Forum. (2024b). *TM forum intent ontology (TIO) (TR292) v3.4.0*. TM Forum Technical Report / Ontology. https://www.tmforum.org/resources/introductory-guide/tm-forum-intent-ontology-tio-v3-4-0-tr292/

W3C. (2014). *RDF 1.1 concepts and abstract syntax*. W3C Recommendation. https://www.w3.org/TR/rdf11-concepts/

W3C. (2017). *Shapes constraint language (SHACL)*. W3C Recommendation. https://www.w3.org/TR/shacl/