

OWL-SDA: A Multi-Agent System for Generating Synthetic Data to Support Ontology Evaluation

Dr. Maxim Van de Wynckel

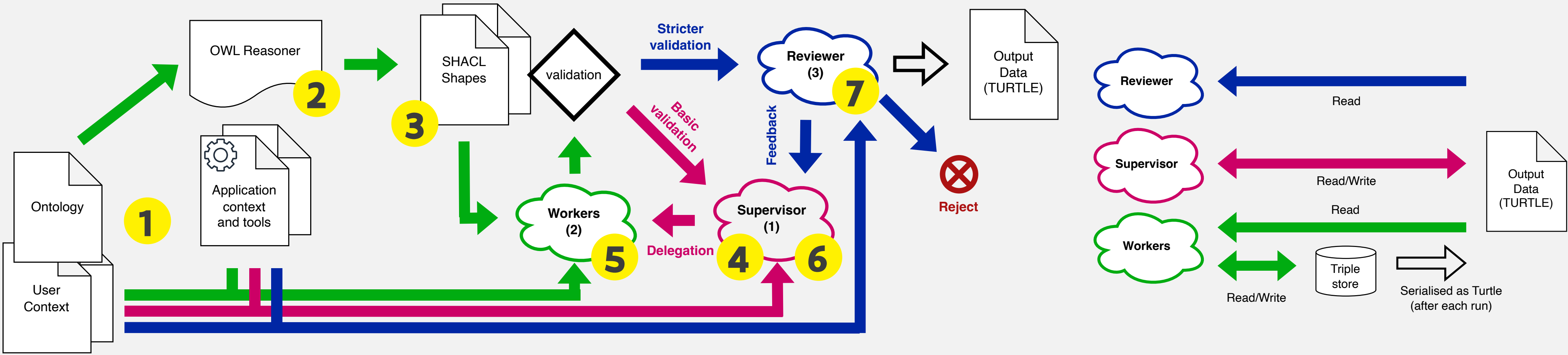
Flemish Government, Department of Environment and Spatial Development, BE
Vrije Universiteit Brussel, Web & Information Systems Engineering Lab, BE
maxim.vandewynckel@vlaanderen.be

Dr. Emmelien De Roock

UZ Brussel & Vrije Universiteit Brussel, Data Analytics Laboratory, BE

SEMANTICS 2026

Ghent, Belgium



Abstract

Ontology engineers often need **representative RDF instances** to validate and document an ontology, but such data may be unavailable, incomplete, too sensitive to share or simply **too much work to create manually**. We present OWL-SDA, an open-source toolkit that uses worker, supervisor and reviewer AI agents to generate synthetic RDF data from ontology context and SHACL constraints. The workflow combines **shape-level task decomposition**, direct manipulation of an in-memory triple store, **iterative SHACL repair** and a final review stage aimed at catching issues that remain possible even after SHACL conformance. We evaluate OWL-SDA on two contrasting ontologies, one for industrial emissions reporting and one for healthcare records. The results show that richer ontology constraints lead to faster convergence and more informative examples, while sparse ontologies require more supervision and still yield less detailed data. We therefore position the toolkit for ontology engineers to **help document their ontology, but also to validate if the ontology alone is sufficiently rich** to be understood by non-experts.

1

Ontology context + user context and application context are provided to the three agents.

2

External ontologies are fetched and a reasoner helps to infer new relations and constraints.

3

Two sets of SHACL shapes are created:
1. Basic SHACL shapes using only the ontology
2. SHACL shapes from the inferred model

4

The supervisor delegates the work to a fixed pool of worker agents. It distributes the workload in tasks (SHACL shapes) sorted by their complexity (shapes with more restrictions first).

5

Workers generate instance data in a shared triple store to facilitate parallelism and automate feedback on the syntax. Workers can validate SHACL conformance using the basic (non-inferred) shapes.

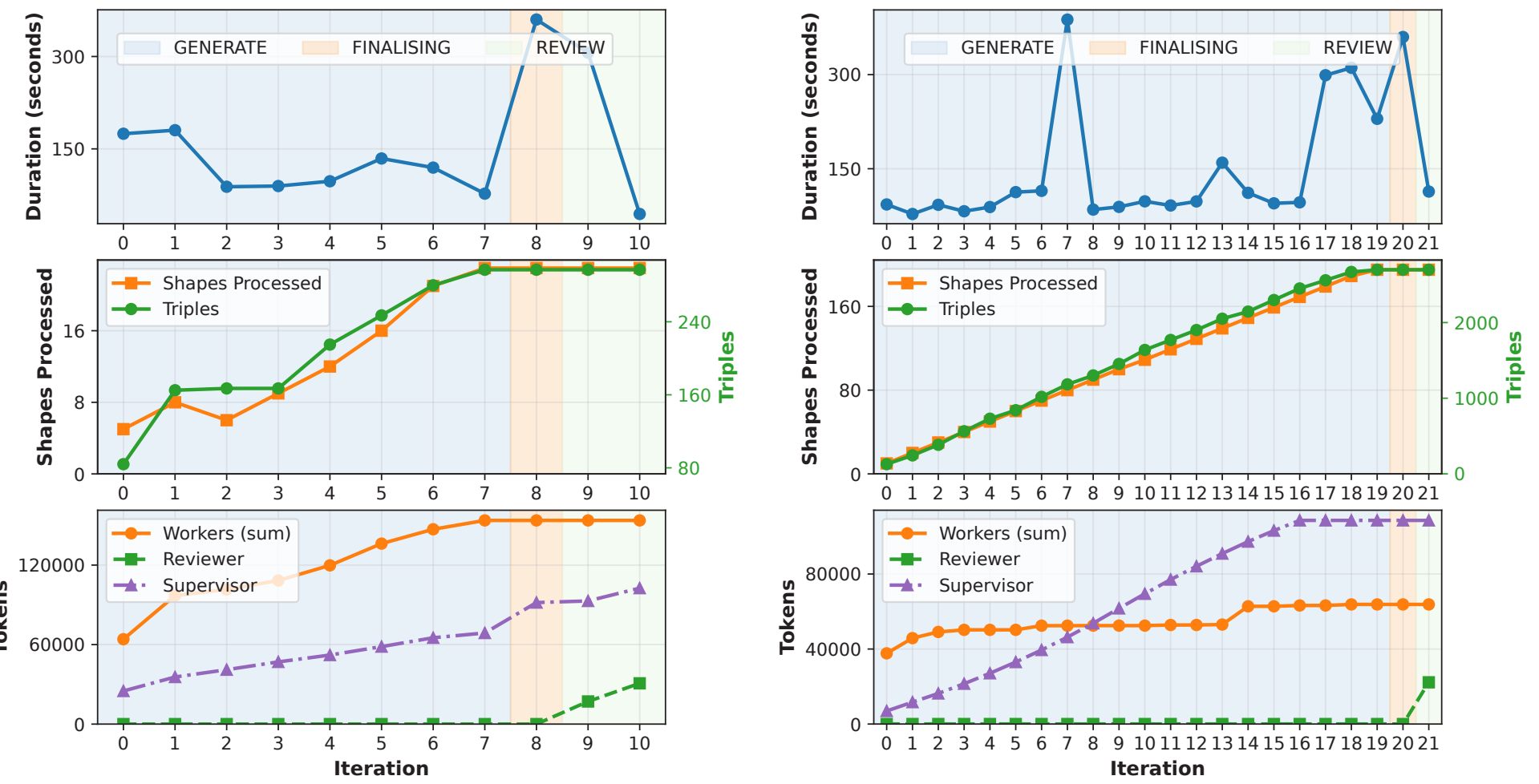
6

The supervisor oversees the final iteration by adding comments and manipulating the file before sending it out for review. The supervisor steers workers to fix inconsistencies and can validate SHACL conformance using the inferred shapes.

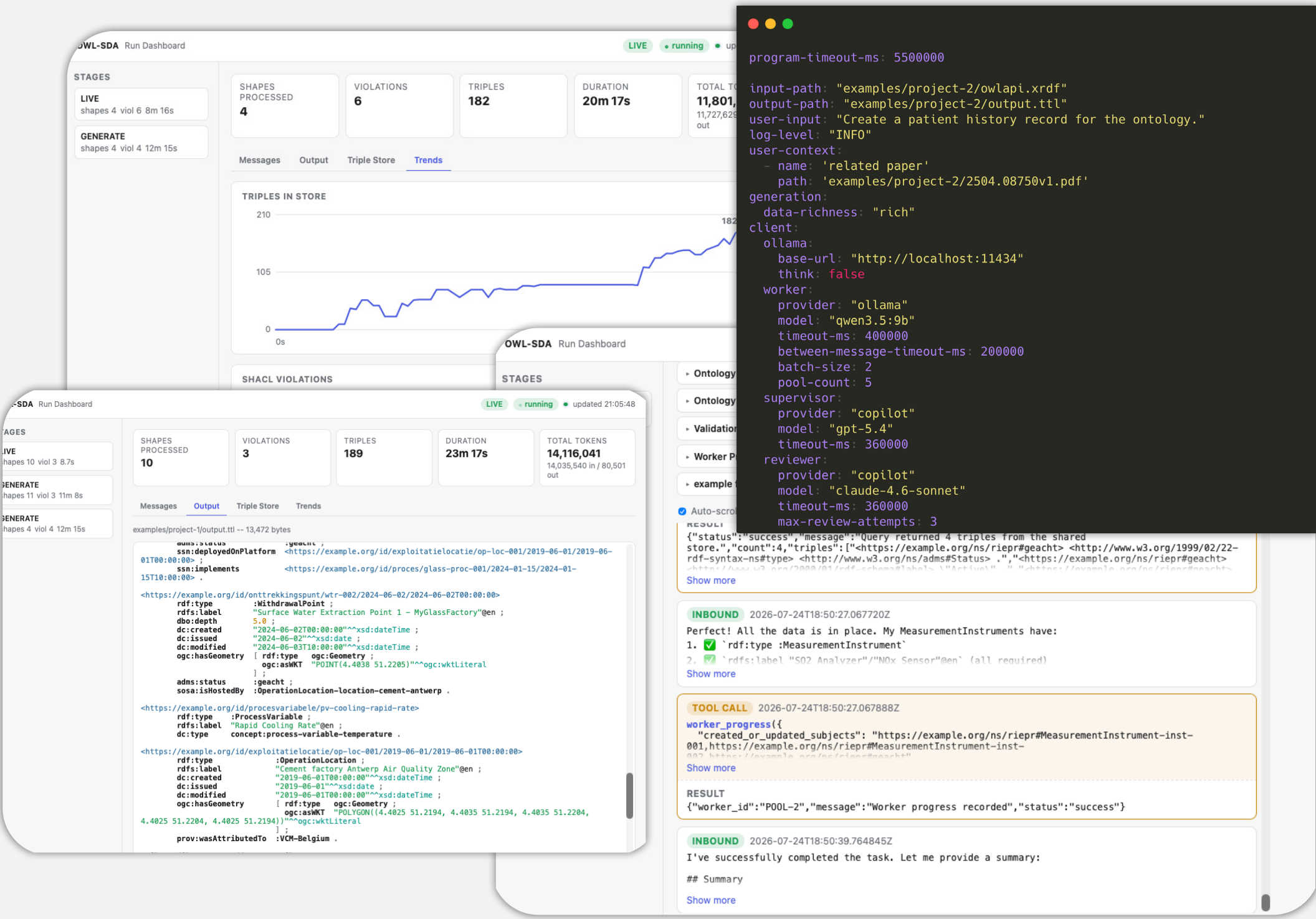
7

The reviewer performs a read-only review on the ontology. Feedback is solved by the supervisor and the review can eventually lead to an accept or reject.

Evaluation



Richer ontology description results in better generated instance data by AI agents. Underspecified ontologies give more freedom to the AI to invent inconsistencies.



Web UI dashboard to monitor agent communication and tool calling

