# Using Knowledge Graphs and SHACL to Validate Declaration Forms: an Experiment in the Social Security Domain to Assess SHACL's Applicability

**Christophe Debruyne[1], Davan Diem Chao[1], Paul Stijfhals[2]**

*[1] Montefiore Institute, University of Liège, Liège, Belgium*
*[2] Smals Research, Smals, Brussels, Belgium*
c.debruyne@uliege.be, davan.chiemdao@student.uliege.be, paul.stijfhals@smals.be

Smals carries out innovative ICT projects in e-government and e-health for social security and health care institutions. Smals Research conducts applied research into novel technologies to see whether they can improve or provide new solutions to their clients. In 2021, these technologies included knowledge graphs and SHACL.

SHACL allows us to validate RDF graphs by declaring constraints called shapes. In public administrations and the EU, SHACL has been used to implement constraints put forward by so-called application profiles. These profiles are used to validate data and are shared with the public. The shapes described in those application profiles are often limited to "simple" integrity constraints such as cardinality, length, and regular expressions.

We wanted to know to what extent SHACL can scale to validate forms against shapes that implement complex business rules, which need to rely on a combination of existing SHACL constructs, logical operations, and embedded SPARQL queries. To answer this question, Smals Research has conducted an experiment in which we validate declaration forms in the Belgian social security (RSZ-ONSS) domain.

A DIMONA declaration is an electronic message with which the employer reports every entry into and exit of an employee to the RSZ-ONSS. One can submit these messages via a Web form, a RESTFUL interface, or via sFTP using an XML Schema developed and maintained by the RSZ-ONSS. The XML schema is a popular tool for bulk uploading such messages, and the schema has integrated simple constraints validating values (cardinality, data type, etc.), much like application profiles. While all constraints are documented, many cannot be tested with XSD and are thus implemented in the application layer and are only validated once uploaded. Some of these constraints refer to so-called "attachments," which contain lists of accepted values and the periods in which they are valid (among others). Another interesting aspect is that the constraints and attachments may evolve every three months, and a declaration has to be validated concerning the constraints applicable at that time.

To answer the question of whether SHACL would scale well and be applicable in such a scenario, we have: 1) created a vocabulary for DIMONA declarations by lifting the conceptual

diagram used for the design of the XML schema, 2) integrated various attachments into a knowledge graph, 3) declared the SHACL constraints that appeared in the documentation (many of which had to be written in SPARQL), and 4) wrote RML mappings to transform DIMONA declarations into RDF for validation.

Our results demonstrate that this approach is viable for validating forms against complex business rules. We show that implementing rules that validate a declaration considering previously submitted declarations is possible, a process that currently relies on application logic. Business rules relying on prior declarations cannot be shared inside an application profile, as that would require access to the knowledge graph, but a web service can be conceived. However, we show that the subset of the shapes one can share constitutes an application profile richer than what is possible with current XSD schemas.

As for future work, we currently investigate knowledge organization strategies for elegantly integrating the ever-evolving documents using named graphs and the integration of validation reports using named graphs and provenance information.

Disclaimer. While Smals builds solutions for the RSZ-ONSS, this study was not conducted for the RSZ-ONSS. This study was conducted to gain insights into the opportunities offered by SHACL. We chose this domain as it was sufficiently representative concerning complex validation rules, and the online availability of all information about DIMONA declarations, the process, and the attachments.