



PGDL: A Interoperable Schema Format for Graphs

Dominik Tomaszuk Łukasz Szeremeta

Institute of Informatics, University of Białystok, Poland

Abstract

Schemes play a key role in databases. They organize data and determine the way in which the database is constructed and what integrity constraints it is affected by. We propose PGDL schema format for Property Graphs, RDF and other graph models.

Introduction

From the two main graph models, RDF and Property graphs, RDF has a rich schema history i.e. Shape Expressions (ShEx)^a, Shapes Constraint Language (SHACL)^b, Resource Shapes (ReSh)^c, Description Set Profiles (DSP)^d, and SPARQL Inferencing Notation (SPIN)^e. On the other hand, Property Graphs have grown in an organic way with every Property Graph Database introducing their own query language including schema parts, i.e. Cypher^f, Gremlin^g, PGQL^h, and GSQLⁱ.

Position Statement

Many existing systems are based on RDF graphs or property graphs. However, there is still no uniform scheme that can be used in each of these systems. At minimum, information exchange about constraints and schema should be helpful for standards work in these two related worlds. The creation of a uniform scheme would also allow for a wider information exchange between graph databases. It may be easier to migrate between different databases as well as to combine data from different sources.

Given the popularity of database systems based on property graphs and RDF, we focus on the schema of data conforming such data models. That is why we propose PGDL that is a interoperable schema format.

^a<http://shex.io/>

^b<https://www.w3.org/TR/shacl/>

^c<https://www.w3.org/Submission/shapes/>

^d<http://dublincore.org/documents/dc-dsp/>

^e<http://spinrdf.org/>

^f<https://www.opencypher.org/>

^g<https://tinkerpop.apache.org/gremlin.html>

^h<http://pgql-lang.org/>

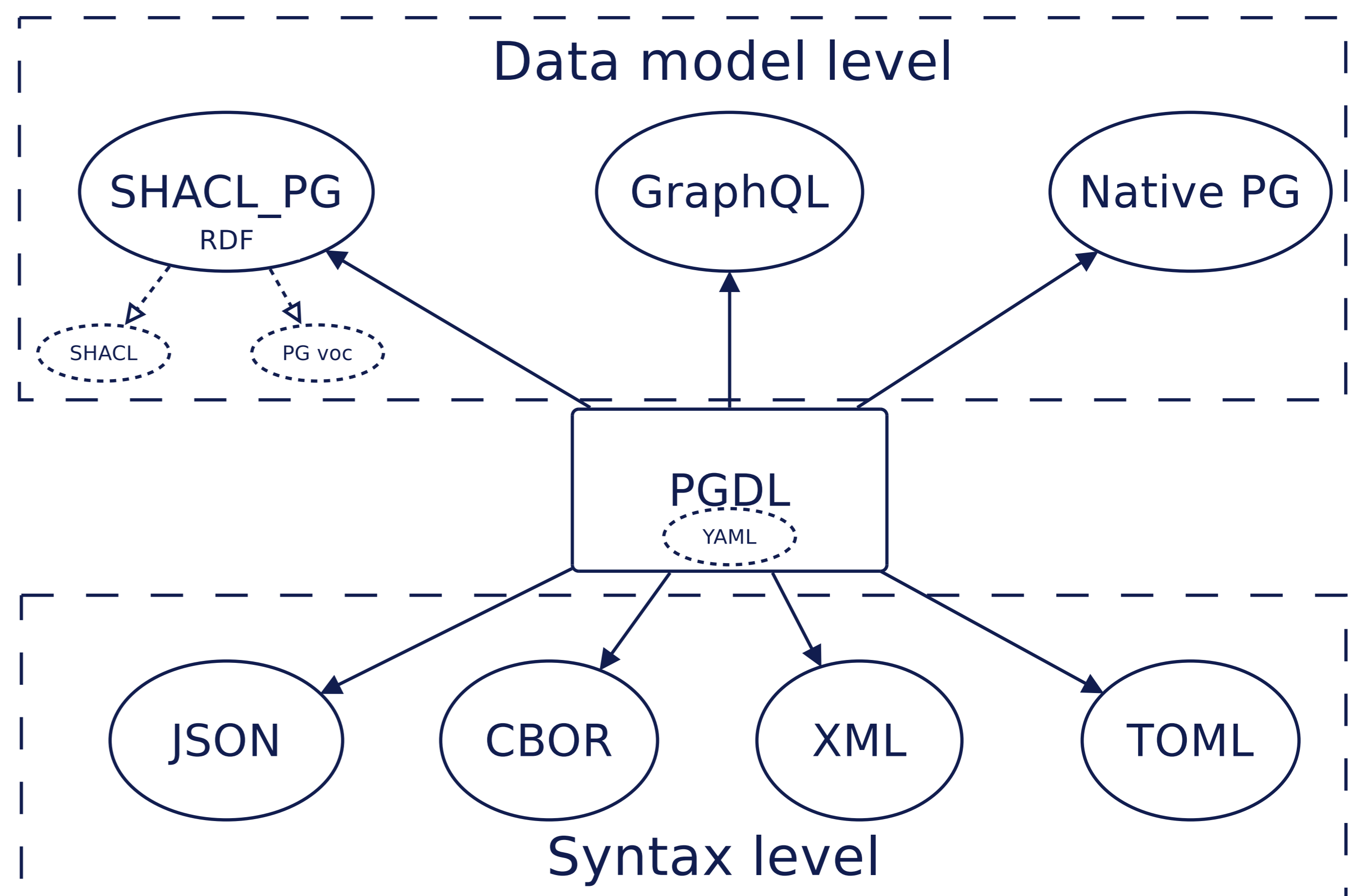
ⁱ<https://doc.tigergraph.com/index.html>

Contact Information

- Dominik Tomaszuk – d.tomaszuk@uwb.edu.pl
- Łukasz Szeremeta – l.szeremeta@uwb.edu.pl

Our proposal

Our proposal supports two levels of interoperability: data model interoperability and schema serialization interoperability. In the level of data model we support RDF [1], Property Graph [2] and GraphQL model [3]. In the level of schema syntax we propose PGDL format based on well-known and human readable YAML [4], that can be transform to additional formats compatible with it, e.g. JSON [5], CBOR [6], XML [7], TOML [8] etc. In addition we have created PGDL-conv (<https://github.com/dome1/PGDL-conv>), an open source command line tool that allows to convert our format to many others, including the ones mentioned above.



Conclusions

Databases allow for checking the conformity of their instances with the given scheme. It is important to provide a format for defining the schema in different data models. We propose PGDL that meets these requirements and support Property Graphs, RDF, and GraphQL. Furthermore, we provide various syntaxes to facilitate the use of our format in most tools and databases.

References

1. RDF – <https://www.w3.org/TR/rdf11-primer/>
2. Dominik Tomaszuk and Łukasz Szeremeta. "Named Property Graphs." 2018 Federated Conference on Computer Science and Information Systems (FedCSIS). IEEE, 2018.
3. GraphQL – <https://facebook.github.io/graphql/>
4. YAML – <https://yaml.org/spec/1.2/spec.html>
5. JSON – <https://tools.ietf.org/html/rfc8259>
6. CBOR – <https://tools.ietf.org/html/rfc7049>
7. XML – <https://www.w3.org/TR/xml11/>
8. TOML – <https://github.com/toml-lang/toml>