

From: d.tomaszuk@uwb.edu.pl
Subject: [Moderator Action] [W3C Workshop] Position statement: Interoperability in graph schemas
Date: 11 January 2019 at 16:52
To: group-data-ws-pc@w3.org



Interoperability in graph schemas

Dominik Tomaszuk [0000-0003-1806-067X], Łukasz Szeremeta [0000-0002-6635-0623]
Institute of Informatics, University of Białystok, Ciołkowskiego 1M, 15-245
Białystok, Poland {d.tomaszuk,[l.szeremeta](mailto:l.szeremeta@uwb.edu.pl)}@uwb.edu.pl

Schemas 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. Databases allow for checking the conformity of their instances with the given scheme.

From the two main graph models, RDF and Property graphs, RDF has a rich schema history i.e. Shape Expressions (ShEx), Shapes Constraint Language (SHACL), Resource Shapes (ReSh), Description Set Profiles (DSP), and SPARQL Inferencing Notation (SPIN). 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, Gremlin, PGQL, and GSQL.

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 our position:

Property Graphs, RDF and other graph models should have a interoperable schema format

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. Performing such operations will not necessarily require broad knowledge of the more specific schemes used in both systems.

Our initial tasks consist of 1) surveying the state of the art, and 2) proposing a new syntax of graph schemas.

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

[1] <https://github.com/domel/PGDL-conv>