

# A Unified Metamodel for NoSQL and Relational Databases\*

Carlos J. Fernández Candel<sup>1</sup>[0000-0002-3835-9428], Diego Sevilla Ruiz<sup>1</sup>[0000-0001-9313-008X], and Jesús García Molina<sup>1</sup>[0000-0003-4685-6659]

Facultad de Informática, Universidad de Murcia, Murcia, España  
{cjferna,dsevilla,jmolina}@um.es

**Keywords:** Unified metamodel, Schema inference, Model-driven Engineering, Schema mappings, NoSQL databases

**Published in:** Information Systems (0306-4379, Inf. Sys.), Volume 104, February 2022, 101898

**Impact Factor:** 2021 JIF=3.18, JCR Category: Information Systems, 83/164 (first position of the third quartile.

**Citation Number:** 28 (excluding 3 self-citations) in Google Scholar.

**Interest for track attendees:** An MDE approach for database engineering is presented, which involves the definition of a metamodel to represent NoSQL and relational systems in a unified way, and bidirectional schema mappings to/from the unified metamodel are specified. Also, algorithms to inject schema models from data stored in schemaless NoSQL systems are described in detail. The paper also discusses the benefits of using MDE techniques in the database engineering scenario. Applications of the metamodel are shown.

**DOI:** <https://doi.org/10.1016/j.is.2021.101898>

**Abstract.** The Database field is undergoing significant changes. Although relational systems are still predominant, the interest in NoSQL systems is continuously increasing. In this scenario, polyglot persistence is envisioned as the database architecture to be prevalent in the future. Therefore, database tools and systems are evolving to support several data models.

Multi-model database tools normally use a generic or unified metamodel to represent schemas of the data model that they support. Such metamodels facilitate developing database utilities, as they can be built on a common representation. Also, the number of mappings required to migrate databases from a data model to another is reduced, and integrability is favored.

In this paper, we present the U-Schema unified metamodel able to represent logical schemas for the four most popular NoSQL paradigms (columnar, document, key-value, and graph) as well as relational schemas. We will formally define the mappings between U-Schema and the data model defined for each database paradigm. How these mappings have been implemented and validated will be discussed, and some applications of U-Schema will be shown.

---

\* This work has been funded by the Spanish Ministry of Science and Innovation (grant project: PID2020-117391GB-I00).



To achieve flexibility to respond to data changes, most of NoSQL systems are “schema-on-read,” and the declaration of schemas is not required. Such an absence of schema declaration makes structural variability possible, i.e., stored data of the same entity type can have different structure. Moreover, data relationships supported by each data model are different; For example, document stores have aggregate objects but not relationship types, whereas graph stores offer the opposite. Throughout the paper, we will show how all these issues have been tackled in our approach.

As far as we know, no proposal exists in the literature of a unified metamodel for relational and the NoSQL paradigms which describes how each individual data model is integrated and mapped. Our metamodel goes beyond the existing proposals by distinguishing entity types and relationship types, representing aggregation and reference relationships, and including the notion of structural variability. Our contributions also include developing schema extraction strategies for schemaless systems of each NoSQL data model, and tackling performance and scalability in the implementation for each store.