

Supporting the Evolution of Wikidata EntitySchemas through Empirical Usage Analysis

Jorge Álvarez-Fidalgo^{1,*}, Enrique Rodríguez-Martin¹, Hugo Lebrede¹ and Jose Emilio Labra-Gayo¹

¹Department of Computer Science, University of Oviedo, Spain

Abstract

Wikidata EntitySchemas provide formal descriptions of the expected structure of entities and play an important role as modelling guidelines and validation artefacts. Existing approaches have primarily focused on schema inference or on assisting editors during data creation through validation mechanisms and property constraints. In contrast, comparatively little attention has been devoted to supporting the maintenance of EntitySchemas once collaborative knowledge graphs have evolved through continuous community contributions.

This paper presents an empirical framework for supporting the evolution of Wikidata EntitySchemas by analysing modelling practices observed in deployed Wikidata data. Starting from a target EntitySchema, the proposed approach constructs an empirical modelling profile from Wikidata instances, capturing property usage, cardinality patterns, qualifier usage, and value distributions. This profile is subsequently contrasted with the formal EntitySchema specification to generate evidence-based maintenance recommendations, including candidate property additions, cardinality updates, qualifier documentation, and value-range refinements.

The proposed approach is demonstrated through a case study on manually curated EntitySchemas from the cultural heritage domain. The obtained results show that empirical usage analysis can identify meaningful maintenance opportunities while preserving the role of EntitySchemas as community-maintained modelling artefacts. By complementing existing schema inference and validation approaches, the proposed framework provides maintainers with quantitative evidence to support the continuous evolution of EntitySchemas as Wikidata modelling practices change over time.

Keywords

Wikidata, EntitySchemas, Shape Expressions, Knowledge Graphs, Schema Evolution, Cultural Heritage

1. Introduction

Wikidata has become one of the largest collaboratively curated knowledge graphs, serving as a fundamental infrastructure for Linked Data, semantic applications, open science, and cultural heritage initiatives. Its open editing model enables thousands of contributors to collaboratively describe entities across a broad range of domains, resulting in a continuously evolving knowledge graph. As modelling practices emerge from community consensus rather than centralised design, the representation of entities naturally changes over time as new properties, qualifiers, and modelling conventions are adopted.

Wikidata provides several complementary mechanisms for promoting modelling consistency and facilitating data validation, including Property Constraints and EntitySchemas. In this work, we focus specifically on *EntitySchemas*, which use Shape Expressions (ShEx) to formally describe the expected structure of entities belonging to a particular class. EntitySchemas are used as modelling guidelines, documentation artefacts, and validation resources, helping editors and applications produce semantically consistent descriptions. Their importance is particularly evident in domains requiring high-quality structured data, such as cultural heritage, scholarly metadata, and biomedical knowledge graphs.

Wikidata'26: Wikidata Workshop at ISWC 2026

*Corresponding author.

✉ alvarezfjorge@uniovi.es (J. Álvarez-Fidalgo); rodriguezmenrique@uniovi.es (E. Rodríguez-Martin);

lebredehugo@uniovi.es (H. Lebrede); labra@uniovi.es (J. E. Labra-Gayo)

🆔 0009-0002-7392-7862 (J. Álvarez-Fidalgo); 0000-0001-7803-8592 (E. Rodríguez-Martin); 0000-0002-8745-2338 (H. Lebrede); 0000-0001-8907-5348 (J. E. Labra-Gayo)



© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

Despite their growing adoption, EntitySchemas remain manually maintained by the community. Existing approaches, including EntitySchemas themselves and Wikidata Property Constraints, primarily support schema authoring or assist editors during data creation and validation. In contrast, much less attention has been devoted to the complementary problem of maintaining EntitySchemas once collaborative knowledge graphs have evolved through years of community contributions.

As Wikidata evolves, modelling conventions naturally change. New properties become widely adopted, qualifiers emerge as de facto modelling conventions, and cardinality assumptions may no longer accurately reflect the way entities are described in practice. These changes do not necessarily indicate modelling errors; rather, they represent the natural evolution of community modelling practices. Consequently, EntitySchemas may require periodic revision to remain representative of the deployed knowledge graph.

Previous research has explored schema inference from knowledge graph instances and the validation of Wikidata entities using formal constraints. However, comparatively little attention has been devoted to supporting the maintenance of existing EntitySchemas through empirical evidence extracted from Wikidata itself. Rather than generating new schemas or validating individual edits, EntitySchema maintainers require evidence indicating whether current modelling practices suggest opportunities for updating existing schemas.

In this paper, we present an empirical framework for supporting the evolution of Wikidata EntitySchemas. Starting from a target EntitySchema, the proposed approach constructs an *empirical modelling profile* from deployed Wikidata instances, capturing property usage, cardinality patterns, qualifier usage, and value distributions. This empirical profile is subsequently contrasted with the corresponding EntitySchema to generate evidence-based maintenance recommendations, including candidate property additions, cardinality updates, qualifier documentation, and value-range refinements.

To demonstrate the feasibility of the approach, we conduct a case study on manually curated EntitySchemas from the cultural heritage domain. This domain provides an appropriate evaluation scenario because it combines active editing communities, heterogeneous modelling practices, and a collection of EntitySchemas maintained by the community. The obtained results show that empirical usage analysis can generate actionable maintenance recommendations while preserving the community-driven evolution of EntitySchemas.

Table 1 positions the proposed framework with respect to representative approaches supporting different stages of the Wikidata schema lifecycle. Unlike existing work on schema inference or validation during editing, our approach focuses on supporting the maintenance of EntitySchemas after the knowledge graph has evolved.

The main contributions of this work are:

- An empirical profiling methodology for characterising modelling practices from collections of Wikidata entities.
- A recommendation framework that contrasts empirical modelling profiles with existing EntitySchemas to generate evidence-based maintenance recommendations.
- A case study on cultural heritage EntitySchemas illustrating how empirical usage analysis can support the evolution of manually curated EntitySchemas.

2. Related Work

The increasing adoption of Wikidata has motivated extensive research on knowledge modelling, schema validation, and the evolution of collaborative knowledge graphs. Existing approaches mainly focus on supporting the creation of modelling artefacts or validating knowledge during editing. Comparatively less attention has been devoted to supporting the evolution of EntitySchemas once collaborative knowledge graphs have matured through continuous community contributions.

Table 1

Representative approaches supporting different stages of the Wikidata schema lifecycle.

Approach	Primary objective	When does it operate?
EntitySchemas	Specify and validate the expected structure of entities	Schema design and validation
Shexer	Automatically generate initial ShEx schemas from instance data	Schema creation
Property Constraints	Validate statements and assist editors during data curation	During data editing
This work	Generate evidence-based recommendations for maintaining EntitySchemas	After knowledge graph evolution

2.1. Shape Expressions and EntitySchemas

Shape Expressions (ShEx) were introduced as a compact language for describing and validating RDF graph structures, providing a formal mechanism for specifying graph constraints while remaining sufficiently expressive for practical knowledge graph applications [1]. Wikidata adopted this formalism through EntitySchemas, which provide machine-readable descriptions of the expected structure of entities belonging to a given class and constitute one of the primary mechanisms for documenting modelling practices and validating Wikidata content.

Several tools have subsequently been proposed to facilitate the creation, management, and validation of EntitySchemas. WShEx extends ShEx to the Wikibase data model, while ShExStatements assists users in generating EntitySchemas from tabular descriptions [2]. Likewise, Wikidata Property Constraints have been formalised using SHACL and SPARQL to provide machine-processable validation rules and editing support, enabling contributors to detect potential inconsistencies during data entry and curation [3].

These approaches primarily support schema authoring and the validation of individual edits before or during knowledge graph curation. In contrast, our framework operates after data has been collaboratively curated, supporting the maintenance of EntitySchemas by analysing the modelling practices that emerge within an already deployed knowledge graph.

2.2. Schema Inference from Knowledge Graphs

A complementary line of research focuses on automatically deriving schemas from existing knowledge graphs. Shexer automatically extracts Shape Expressions from RDF graphs by identifying recurring structural patterns, providing an effective mechanism for generating initial schema specifications from instance collections [4].

More recently, Carriero et al. proposed extracting empirical ontology design patterns from Wikidata and transforming them into probabilistic Shape Expressions, demonstrating that recurring modelling conventions can be automatically identified from large collections of entities and compared with existing modelling artefacts [5]. Automatic schema generation has also been explored using large language models capable of synthesising ShEx schemas directly from knowledge graph content [6].

Unlike these approaches, our objective is not to synthesise new EntitySchemas from existing data. Instead, we assume that manually curated EntitySchemas already exist and investigate whether empirical modelling behaviour observed in deployed Wikidata data suggests opportunities for their evolution. In this sense, empirical evidence is used to support schema maintenance rather than schema generation.

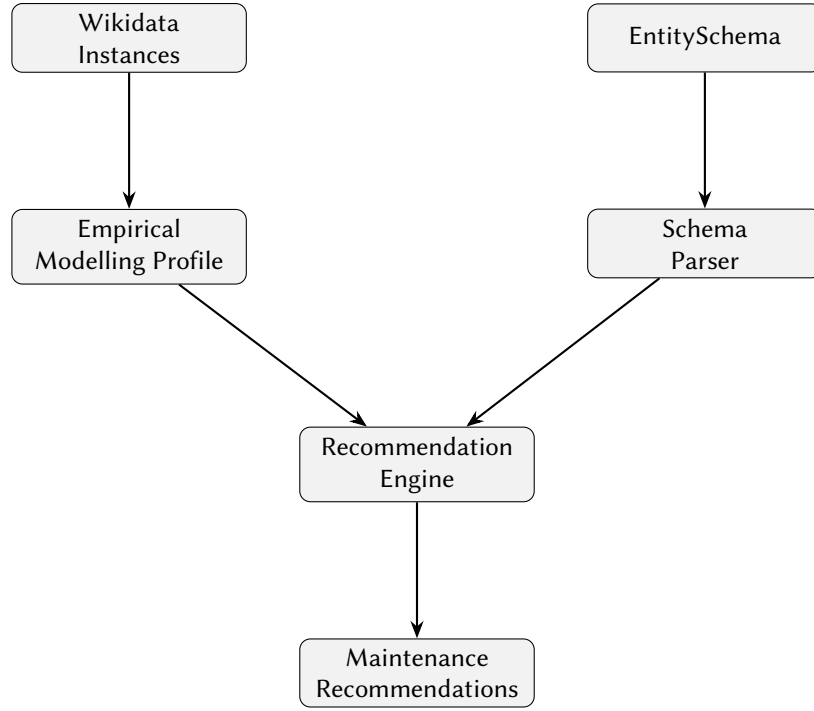


Figure 1: Overview of the proposed recommendation framework. An empirical modelling profile extracted from Wikidata instances is compared with the formal EntitySchema specification to generate evidence-based maintenance recommendations.

2.3. Evolution of Collaborative Knowledge Graphs

Wikidata has been extensively studied as a collaborative knowledge graph whose content continuously evolves through community contributions. Previous research has investigated aspects such as data completeness, quality assessment, collaborative editing dynamics, and the evolution of modelling practices, showing that modelling conventions naturally change over time as new entities, properties, qualifiers, and editing strategies emerge [7].

Although these studies characterise how Wikidata evolves, they do not provide automated mechanisms for translating empirical modelling behaviour into actionable recommendations for maintaining EntitySchemas. Consequently, maintainers still lack systematic support for determining whether existing EntitySchemas remain representative of the modelling conventions currently adopted by the community.

The framework presented in this paper addresses this gap by constructing empirical modelling profiles from deployed Wikidata entities and contrasting them with existing EntitySchemas. Rather than replacing schema engineering or validating individual edits, the proposed approach generates evidence-based maintenance recommendations that assist maintainers in evolving EntitySchemas as collaborative modelling practices change over time.

3. Methodology

The proposed framework supports the maintenance of Wikidata EntitySchemas by contrasting their formal specification with the modelling practices observed in Wikidata. Rather than validating individual entities, the objective is to characterise how a community models a particular class of entities and to determine whether the corresponding EntitySchema accurately reflects these practices.

Figure 1 summarises the proposed workflow. Starting from a target EntitySchema, the framework retrieves a collection of Wikidata instances belonging to the corresponding class, constructs an empirical modelling profile, analyses the EntitySchema specification, and finally generates evidence-based maintenance recommendations by comparing both representations.

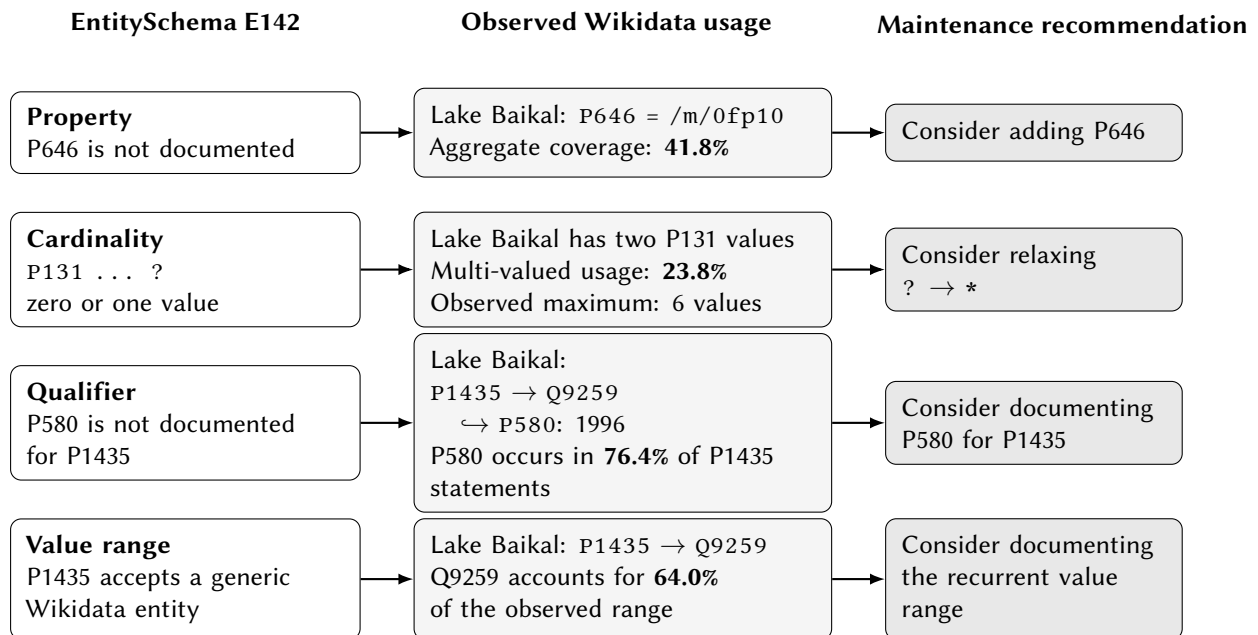


Figure 2: Running example illustrating the four types of maintenance evidence considered by the framework for the UNESCO World Heritage Site EntitySchema (E142). Lake Baikal (Q5513) exemplifies individual modelling patterns, whereas maintenance recommendations are generated from their aggregated prevalence across the analysed entities.

Running example. To illustrate the concepts used throughout the methodology, we consider the UNESCO *World Heritage Site* EntitySchema (E142) and *Lake Baikal* (Q5513), one of its corresponding Wikidata entities. EntitySchemas use Shape Expressions (ShEx) to describe expected modelling patterns. For example, E142 declares *located in the administrative territorial entity* (P131) with optional single-value cardinality (?) and *heritage designation* (P1435) with zero-or-more cardinality (*):

```
wdt:P131 [ <http://www.wikidata.org/entity>~ ] ? ;
wdt:P1435 [ <http://www.wikidata.org/entity>~ ] * ;
```

Lake Baikal illustrates how deployed Wikidata data may exhibit modelling patterns that go beyond these schema declarations. It contains two values for P131, uses *Freebase ID* (P646), which is absent from E142, and represents P1435 as a statement whose value is *World Heritage Site* (Q9259) and which is qualified by *start time* (P580).

These observations correspond to the four dimensions considered by our framework: property usage, cardinality, qualifier usage, and value ranges. Figure 2 summarises the running example by contrasting the current EntitySchema with observed Wikidata usage and the corresponding type of maintenance recommendation.

Importantly, individual entities do not directly trigger recommendations. The framework aggregates these observations across the analysed collection and generates recommendations only when recurrent patterns satisfy the empirical support criteria described below.

3.1. Empirical Modelling Profile Construction

Given an EntitySchema associated with a Wikidata class, the first step consists of retrieving a collection of entities belonging to that class through the Wikidata SPARQL endpoint. The corresponding entities are then downloaded using the Wikidata API, preserving the complete statement structure, including qualifiers and references.

The downloaded entities are analysed to construct an *empirical modelling profile*, which summarises the modelling practices observed across the dataset. For every property appearing in the analysed entities, the framework computes several descriptive statistics, including:

- entity coverage, defined as the proportion of entities containing the property;
- observed cardinality statistics, including minimum, maximum and average number of values per entity;
- proportion of entities using multiple values for the property;
- qualifier usage frequencies associated with each property;
- empirical distributions of observed values for entity-valued properties.

The resulting profile provides a compact statistical description of how the community models entities belonging to the analysed class, independently of any existing EntitySchema.

3.2. EntitySchema Analysis

The corresponding EntitySchema is retrieved from Wikidata and parsed into an internal representation suitable for analysis. The parser extracts all structural constraints explicitly declared in the schema, including:

- documented properties;
- cardinality constraints;
- qualifier constraints;
- value restrictions and class constraints.

This representation constitutes the formal description of the intended modelling conventions against which empirical observations are compared.

3.3. Recommendation Generation

The empirical modelling profile and the parsed EntitySchema are subsequently compared to identify opportunities for schema maintenance. Instead of detecting modelling errors, the framework generates recommendations whenever empirical evidence suggests that the documented schema no longer fully reflects current community practices.

Four categories of recommendations are currently supported, as illustrated by the running example in Figure 2.

Property recommendations. Frequently used properties that are not documented in the EntitySchema are proposed as candidate additions. Recommendations are generated only when the empirical coverage of the property exceeds configurable thresholds, distinguishing between frequently used and core modelling properties.

Cardinality recommendations. Observed cardinalities are compared with those declared in the EntitySchema. Recommendations are generated only when a substantial proportion of entities use multiple values for a property, preventing isolated modelling exceptions from producing unnecessary suggestions.

Qualifier recommendations. Qualifier usage is analysed independently for every property. When a qualifier consistently appears in statements of a given property but is not documented in the corresponding EntitySchema, the framework recommends documenting its usage.

Value range recommendations. For entity-valued properties, the framework analyses the empirical distribution of observed values. Frequently occurring values that are not explicitly represented in the EntitySchema are proposed as candidate value constraints or refinements.

3.4. Recommendation Ranking

Each recommendation is accompanied by the empirical evidence supporting it, including coverage statistics, observed cardinalities, qualifier frequencies or value distributions, depending on the recommendation type.

Recommendations are assigned a confidence level according to configurable empirical criteria specific to each recommendation type. These confidence levels are used to rank the generated recommendations according to the strength of the supporting empirical evidence.

The final output of the framework is therefore not a validation report, but a ranked collection of evidence-based maintenance recommendations intended to assist EntitySchema maintainers during the evolution of Wikidata modelling practices.

4. Case Study

To evaluate the proposed framework, we conducted a case study on a selected subset of Wikidata EntitySchemas from the cultural heritage domain. This domain was selected for three main reasons. First, cultural heritage constitutes one of the most active modelling communities within Wikidata, where entities are collaboratively curated by contributors from different countries and institutions. Second, the domain exhibits heterogeneous modelling practices involving geographical, administrative, historical, and bibliographic information, making it particularly suitable for studying the evolution of modelling conventions. Finally, several manually curated EntitySchemas are already available, providing realistic maintenance scenarios.

4.1. Selected EntitySchemas

Three EntitySchemas were selected to represent different categories of cultural heritage resources:

- **E125 – Museum**
- **E142 – UNESCO World Heritage Site**
- **E476 – Manuscript**

These schemas differ in both their complexity and the number of Wikidata entities they describe, allowing the proposed framework to be evaluated across modelling scenarios of different sizes and characteristics.

4.2. Experimental Configuration

For each EntitySchema, the corresponding Wikidata class was identified and used to retrieve a collection of entities through the Wikidata SPARQL endpoint. Entity descriptions were subsequently downloaded using the Wikidata API, preserving complete statement structures, including qualifiers.

The proposed framework was then applied independently to each EntitySchema following the workflow described in Section 3. For every case study, an empirical modelling profile was constructed from the collected entities and compared with the formal EntitySchema specification to generate maintenance recommendations.

Recommendations were produced according to a set of configurable empirical thresholds. Properties appearing in at least 30% of the analysed entities were considered *frequent properties*, while properties appearing in at least 80% of the entities were considered *core properties*. Candidate cardinality updates were generated only for properties appearing in at least 20% of the entities and whose multi-value usage ratio was at least 20%, preventing isolated modelling exceptions from producing recommendations. Qualifier and value-range recommendations were generated when the corresponding empirical evidence reached a support threshold of at least 50%.

Table 2

Summary of the empirical modelling profiles.

Schema	Observed properties	Frequent properties	Core properties	Frequent recall	Core recall
Museum	1540	12	4	83.3%	100.0%
World Heritage Site	436	12	3	83.3%	100.0%
Manuscript	504	8	1	62.5%	100.0%

4.3. Evaluation Strategy

The objective of this study is not to evaluate the correctness of EntitySchemas, but to assess whether empirical modelling behaviour can be exploited to generate meaningful maintenance recommendations. Consequently, the evaluation focuses on characterising the recommendations produced by the framework and the empirical evidence supporting them.

We additionally report frequent- and core-property recall. Frequent- and core-property recall are then defined as the proportion of properties in each category that are already documented in the corresponding EntitySchema. These measures characterise the extent to which existing EntitySchemas cover recurrent modelling practices and should not be interpreted as precision or recall of the generated recommendations.

No gold standard of EntitySchema maintenance decisions is currently available. Therefore, the present evaluation does not estimate recommendation precision or correctness against a reference set. Instead, we analyse whether the generated recommendations correspond to modelling patterns that are consistently observed in the analysed Wikidata entities. Particular attention is paid to four recommendation categories: property additions, cardinality updates, qualifier documentation, and value-range refinements.

5. Results

Table 2 summarises the empirical modelling profiles generated for the analysed EntitySchemas. Although the selected schemas belong to the same application domain, they exhibit considerable differences in size and modelling complexity. The number of observed properties substantially exceeds the number explicitly documented in the corresponding EntitySchemas, reflecting the richness of collaborative modelling practices in Wikidata.

Despite this variability, the analysed EntitySchemas capture a substantial proportion of the recurrent modelling conventions identified by the empirical profiles. In all cases, the recall of core properties reaches 100%, while recall over frequently used properties exceeds 60%, suggesting that manually curated EntitySchemas effectively document the core characteristics of the represented classes. This suggests that EntitySchemas primarily capture the defining modelling characteristics of a class rather than attempting to exhaustively document every property observed in practice.

5.1. Generated Maintenance Recommendations

Applying the proposed framework produced four categories of maintenance recommendations: candidate property additions, cardinality updates, qualifier documentation, and value-range refinements. Their distribution varied across the analysed EntitySchemas, reflecting differences in community modelling practices rather than deficiencies in the schemas themselves (Table 3). Qualifier documentation clearly dominates the generated recommendations, suggesting that EntitySchemas generally capture the principal properties of a class but often lag behind the increasingly rich qualifier usage adopted by the Wikidata community.

Table 3

Distribution of generated maintenance recommendations.

Schema	Properties	Cardinality	Qualifiers	Ranges	Total
Museum	2	0	15	0	17
World Heritage Site	2	1	2	1	6
Manuscript	3	0	1	2	6

5.2. Representative Recommendations

Several recommendations generated by the framework illustrate how empirical modelling evidence can support the evolution of EntitySchemas.

The most frequently generated recommendations concern *qualifier documentation*. Across the analysed EntitySchemas, several qualifiers consistently appear in statements but are not explicitly documented in the corresponding schemas. For example, for the property *heritage designation* (P1435), the qualifier *start time* (P580) appears in the majority of observed statements but is absent from the analysed EntitySchema. Rather than indicating an error, this recommendation suggests documenting a modelling convention that has become widely adopted by the community.

The second most common recommendation concerns *candidate property additions*. The framework identifies properties that are consistently used across a substantial proportion of the analysed entities but are not currently documented in the EntitySchema. For instance, properties such as *has parts* (P527) in the World Heritage Site EntitySchema or *Freebase ID* (P646) in all analysed schemas satisfy the empirical support thresholds and are therefore proposed as candidates for inclusion. Since recommendations are generated only for frequently occurring properties, isolated modelling practices do not trigger unnecessary schema modifications.

Recommendations concerning *value-range refinements* were also identified for several properties. These arise when the observed values of an entity-valued property consistently belong to a particular Wikidata class that is not explicitly documented in the corresponding EntitySchema. For instance, in the World Heritage Site EntitySchema, *World Heritage Site* (Q9259) accounts for 64% of the observed values of the property *heritage designation* (P1435), suggesting a recurrent modelling pattern that is absent from the current schema. Rather than inferring a new restriction automatically, the framework recommends documenting this empirical pattern so that maintainers can decide whether it should become part of the formal EntitySchema.

Finally, only a single *cardinality recommendation* was generated across the analysed case studies. For the property *located in the administrative territorial entity* (P131) in the World Heritage Site EntitySchema, more than 20% of the entities using this property contain multiple values, suggesting that the documented optional single-value cardinality could be relaxed. The relatively small number of cardinality recommendations reflects the conservative thresholds adopted by the framework, which intentionally avoids proposing schema modifications based on isolated modelling exceptions.

6. Conclusions

This paper presented an empirical framework for supporting the evolution of Wikidata EntitySchemas through the analysis of deployed Wikidata data. Rather than inferring new schemas or validating individual edits, the proposed approach constructs empirical modelling profiles and contrasts them with existing EntitySchemas to generate evidence-based maintenance recommendations.

The case study suggests that manually curated EntitySchemas generally capture the most recurrent modelling conventions of their corresponding classes while also revealing opportunities for incremental evolution, particularly regarding qualifier documentation. Across the analysed schemas, the most common recommendations concerned qualifier documentation, followed by candidate property additions and value-range refinements, whereas cardinality updates were comparatively rare. These findings suggest that empirical modelling profiles provide useful evidence for assisting EntitySchema

maintainers without replacing the community-driven schema engineering process.

The proposed framework is intended as a decision-support tool rather than an automatic schema generation system. Since recommendations are derived exclusively from empirical modelling behaviour, they should be interpreted as evidence to assist maintainers, who ultimately decide whether observed modelling conventions should be incorporated into the corresponding EntitySchemas.

A limitation of the approach is that empirical prevalence does not necessarily imply semantic correctness. Systematic modelling errors may therefore produce recommendations when they are sufficiently widespread, while semantically important properties affected by substantial missingness may fail to reach the configured support thresholds. Moreover, the current framework does not model dependencies introduced by shared editors, bots, or copy-based editing practices. The generated recommendations should consequently be interpreted as empirical evidence for maintainer review rather than as automatically valid schema modifications.

Future work will extend the evaluation to a broader range of Wikidata domains and involve experienced Wikidata editors in assessing the usefulness of the generated recommendations. We also plan to investigate the sensitivity of the recommendations to different empirical threshold configurations and to compare the proposed maintenance-oriented approach with methods that generate schemas from scratch. Wikidata’s revision history further provides an opportunity for retrospective evaluation by comparing recommendations generated from historical snapshots with subsequent manual changes to EntitySchemas. Another extension would incorporate Wikidata’s class hierarchy when analysing entity-valued properties, allowing value-range recommendations to consider subclass relations and transitive P279 paths rather than only directly observed values. Finally, we will explore integration with Wikimedia infrastructure and user-facing maintenance workflows through which recommendations can be directly inspected and reviewed by EntitySchema maintainers.

Availability

The implementation of the proposed framework, together with example configurations and the data used in the case study, is publicly available at:

<https://github.com/ExarcaFidalgo/wikidata-entityschema-maintainer>

Acknowledgments

This work was partially funded by the Spanish Research Agency project “Dando forma a los grafos de conocimiento e interoperables” (PID2024-157010OB-I00), and by the regional project SEK-25-GRU-GIC-24-089.

Declaration on Generative AI

During the preparation of this work, the authors used ChatGPT-5 in order to: Grammar and spelling check. After using these services, the authors reviewed and edited the content as needed and take full responsibility for the publication’s content.

References

- [1] E. Prud’hommeaux, J. E. Labra Gayo, H. Solbrig, Shape expressions: an rdf validation and transformation language, in: Proceedings of the 10th International Conference on Semantic Systems, 2014.
- [2] J. E. L. Gayo, Wshex: A language to describe and validate wikibase entities, 2022. URL: <https://arxiv.org/abs/2208.02697>. arXiv: 2208.02697.

- [3] N. Ferranti, J. F. De Souza, S. Ahmetaj, A. Polleres, Formalizing and validating wikidata's property constraints using shacl and sparql, *Semantic Web* 15 (2024) 2333–2380. doi:10.3233/SW-243611.
- [4] D. Fernández-Álvarez, J. E. Labra-Gayo, D. Gayo-Avello, Automatic extraction of shapes using shexer, *Knowledge-Based Systems* 238 (2022) 107975. doi:10.1016/j.knosys.2021.107975.
- [5] V. A. Carriero, P. Groth, V. Presutti, Empirical ontology design patterns and shapes from wikidata, *Semantic Web* 15 (2024) 2293–2317. URL: <https://journals.sagepub.com/doi/abs/10.3233/SW-243613>. doi:10.3233/SW-243613. arXiv:<https://journals.sagepub.com/doi/pdf/10.3233/SW-243613>.
- [6] B. Zhang, Y. He, L. Pintscher, A. Meroño-Peñuela, E. Simperl, Schema generation for large knowledge graphs using large language models, in: C. Christodoulopoulos, T. Chakraborty, C. Rose, V. Peng (Eds.), *Findings of the Association for Computational Linguistics: EMNLP 2025*, Association for Computational Linguistics, Suzhou, China, 2025, pp. 12561–12580. URL: <https://aclanthology.org/2025.findings-emnlp.671/>. doi:10.18653/v1/2025.findings-emnlp.671.
- [7] M. Luggen, D. Difallah, C. Sarasua, G. Demartini, P. Cudré-Mauroux, *Non-parametric Class Completeness Estimators for Collaborative Knowledge Graphs—The Case of Wikidata*, Springer International Publishing, 2019, p. 453–469. URL: http://dx.doi.org/10.1007/978-3-030-30793-6_26. doi:10.1007/978-3-030-30793-6_26.