

Enforcing fine-grained data governance for Wikibase via generic and declarative filter rules

Sven Lieber^{1,*}

¹KBR - Royal Library of Belgium, Keizerslaan 4, 1000 Brussels, Belgium

Abstract

The MetaBelgica consortium — four Belgian Federal Scientific Institutions led by KBR, the Royal Library of Belgium — collaboratively curates shared authority data about authors and artists connecting its cultural heritage collections. Because this data partially includes personal information, its publication must follow the principle “as open as possible, as closed as necessary”. We adopted Wikibase as a mature platform for collaborative curation by non-technical users, but its access model is effectively all-or-nothing at the instance level, with no item- or property-level granularity. This poster explores alternative approaches for modelling fine-grained governance in Wikibase, ranging from statement-level annotations to explicit governance entities. We discuss their respective trade-offs in terms of curator effort, consistency, and expressiveness. To support the discussion, we present a work-in-progress filter API that implements one of these approaches and enables controlled publication of Wikibase content. We seek feedback on how governance rules should be represented in Wikibase-based knowledge graphs before validating the proposed approach in the MetaBelgica production environment.

Keywords

Wikibase, Data governance, Privacy, GDPR

1. Introduction

The MetaBelgica consortium brings together four Belgian Federal Scientific Institutions that collaboratively curate shared authority data — primarily persons such as authors and artists — which connects our otherwise heterogeneous cultural heritage collections. Because this data partially includes personal information, its publication is governed by the principle “as open as possible, as closed as necessary”.

To replace legacy systems that lack collaborative curation and de-duplication features, we adopted a centralized Wikibase instance [1]: a mature platform that enables non-technical curators to directly maintain a shared knowledge graph [2], with established advantages for research data management compared to the public Wikidata instance, such as curating original research data, applying various licenses, and the option to remain closed to the open web [3].

However, Wikibase’s access model is effectively all-or-nothing at the instance level, with no item-level or property-level granularity — a mismatch with our governing principle that grows sharper as GDPR compliance and automated agents on the web raise the stakes of fine-grained data governance in terms of what is made accessible.

This work explores a design space of alternative approaches for fine-grained governance in Wikibase. We seek feedback on suitable modelling strategies that balance curator effort, consistency, and expressiveness. At this stage, we complement this exploration with a first prototype that illustrates one of these approaches and supports early feedback from the community; a detailed evaluation is left for future work.

Posters, Demos, Blue Sky, and Tutorials at SEMANTiCS 2026, Sep 2026, Ghent, Belgium

*Corresponding author.

✉ sven.lieber@kbr.be (S. Lieber)

🌐 <https://sven-lieber.org/en> (S. Lieber)

🆔 0000-0002-7304-3787 (S. Lieber)



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

2. Related Work

Fine-grained access control and policy-aware data publication have been widely studied in Semantic Web, RDF infrastructures, and knowledge graph systems. In particular, vocabularies such as the W3C Data Privacy Vocabulary (DPV) [4], PROV-O [5] for provenance tracking, and the ODRL model [6] for expressing usage policies provide formal mechanisms to describe permissions, obligations, and consent in machine-readable form. However, these models are typically applied at the level of open or interoperable RDF datasets rather than integrated into curator-facing tools for Wikibase-based knowledge graph editing.

Within the Wikibase ecosystem, existing work and tooling primarily target open instances and rely on SPARQL endpoints for querying and exploration. Tools such as Reasonator¹ and Scholia² generate entity-centric views by executing SPARQL queries over publicly accessible data, often introducing latency due to query execution and aggregation. SQUID³ extends this idea towards structured exploration of classes and properties, supporting Wikidata editors in data curation tasks, but remains dependent on openly accessible endpoints.

The *sta-doc*⁴ project at the German National Library, uses a Wikibase to store data curation rules and generate documentation. It stores openly available curation rules and does not define governance on data.

Several application-specific systems build additional interfaces on top of Wikibase. The *Resistance in Belgium* project⁵ provides a search interface over Wikibase content, while tools such as *Cradle*⁶ support form-based data entry. However, these systems assume publicly available data or operate within traditional read/write workflows, rather than addressing fine-grained governance or restricted publication.

A key limitation of SPARQL-based and UI-driven approaches is their dependence on openly accessible endpoints and infrastructure components such as the Wikibase Query Service (WDQS). In private Wikibase deployments, this becomes problematic because synchronization components (e.g., the WDQS updater) do not support authentication, preventing their use in restricted environments.

More recent work on Wikibase access control approaches highlights emerging patterns for representing policies directly in Wikibase statements, including reusable policy entities attached to statements [7]. While this demonstrates increasing interest, it still leaves open how such policy representations can be made practical for curator-driven modelling and integration into lightweight access layers.

Overall, while semantic policy vocabularies and Wikibase tooling exist in parallel, there remains a gap between formal policy models and lightweight, curator-facing mechanisms for enforcing fine-grained governance in private Wikibase instances. This motivates our investigation of alternative modelling approaches for integrating governance rules into Wikibase-based workflows.

3. MetaBelgica use case

The MetaBelgica consortium consists of four Belgian Federal Scientific Institutions – the Royal Library of Belgium (KBR), the Royal Museums of Fine Arts of Belgium, the Royal Institute for Cultural Heritage, and the Royal Museums of Art and History – that collaboratively curate shared authority data linking their cultural heritage collections.

From more than one million entities, the dataset contains over 960,000 person records, making personal data governance a central concern, particularly for living persons. As KBR is responsible for the Belgian legal deposit, a large portion of these records relates to contemporary authors, hence possibly living persons. While certain attributes are required for entity disambiguation, the consortium

¹<https://meta.wikimedia.org/wiki/Reasonator>

²<https://scholia.toolforge.org/>

³<https://sqid.toolforge.org/>

⁴<https://github.com/deutsche-nationalbibliothek/sta-doc>

⁵<https://resistanceinbelgium.be/>

⁶<https://github.com/magnusmanske/cradle>

follows the data minimization principle and aims to avoid unnecessary disclosure. This also includes the need to handle potential GDPR requests by data subjects (the authors/artists).

Curation is performed by non-technical domain experts who jointly maintain and de-duplicate entities in a shared environment. To support this collaboration, the consortium adopted a centralized Wikibase instance as a lightweight knowledge graph platform for semantic curation [1]. However, shared editing combined with partially sensitive data introduces the need for selective visibility mechanisms not provided by the default system.

A key requirement of the MetaBelgica use case is that governance must remain accessible to non-technical curators while supporting fine-grained control at both item and property level. Governance rules must also allow temporal validity, support legal justifications such as consent or opt-out, and remain compatible with standards such as ODRL, DPV, and PROV-O. At the same time, the system must avoid excessive modelling complexity that would hinder day-to-day curation workflows.

Although the initial MetaBelgica project concludes in December 2026, collaboration between the institutions will continue, with a production environment that was procured via public tender in the summer of 2026. The governance requirements discussed in this work are therefore not limited to a prototype setting but must support long-term operational use.

4. Approach

This section presents our approach to declaratively modelling data access rules using Wikibase items and properties. This offers two main advantages. First, non-technical curators can create and modify access rules through the existing Wikibase user interface. Second, all changes to these rules are automatically tracked through Wikibase’s built-in revision history, making them auditable. There are different ways to model the rules with Wikibase functionality, hence we present several alternative modelling approaches for fine-grained governance in Wikibase. The following section discusses a prototype implementation that realizes one of these approaches.

Statement-level annotations. A first option is to attach visibility information directly to individual statements using qualifiers or to items/properties (see fig. 1, top left). For example, the property *P19* (place of birth) can be annotated with *visibility = public* as a default setting, while a specific statement such as *Q23 → P19* (place of birth) can override this default via a qualifier setting *visibility = private*. This approach is closely aligned with the native Wikibase data model and is intuitive for curators, since governance metadata is colocated with the governed data. It can be interpreted as a lightweight, implicit access-control policy, loosely corresponding to an ODRL-like rule attached to a statement, while DPV annotations such as *dpv:hasLegalBasis* can be added as additional qualifiers on the same statement (fig. 1, (a)). However, richer governance semantics such as temporal validity or provenance remain implicit and require additional modelling conventions, and governance decisions are not represented as first-class objects. Additionally, any data-access related qualifier actually qualifies the property which is not semantically sound, i.e. it is not the birth place Ghent that has a legal basis, but the visibility statement ⁷.

Centralized visibility property. A second approach introduces a dedicated visibility property at the item level, combined with qualifiers that indicate which property the rule applies to (fig. 1, bottom left). In this model, *Q23* may contain *visibility = private* with qualifier *appliesTo P19 (place of birth)*, and similar visibility statements can be repeated for other properties. This reduces redundancy and centralises governance decisions per item, while still allowing property-level control. Temporal validity can be modelled directly on the visibility statement using start and end time qualifiers, while legal justification can be attached uniformly (e.g. via *dpv:hasLegalBasis*). In ODRL terms, this corresponds more closely to a rule structure where a rule combines an action (e.g. disclosure or access) with a target scope (i.e. the affected property or item). However, this approach can be prone to error, in particular when specifying the target of the rule (fig. 1, (b)), one could specify a property that does

⁷Please note that natively a Wikibase does not adhere to *formal semantics* like an OWL ontology, yet, qualifiers are meant to qualify a statement

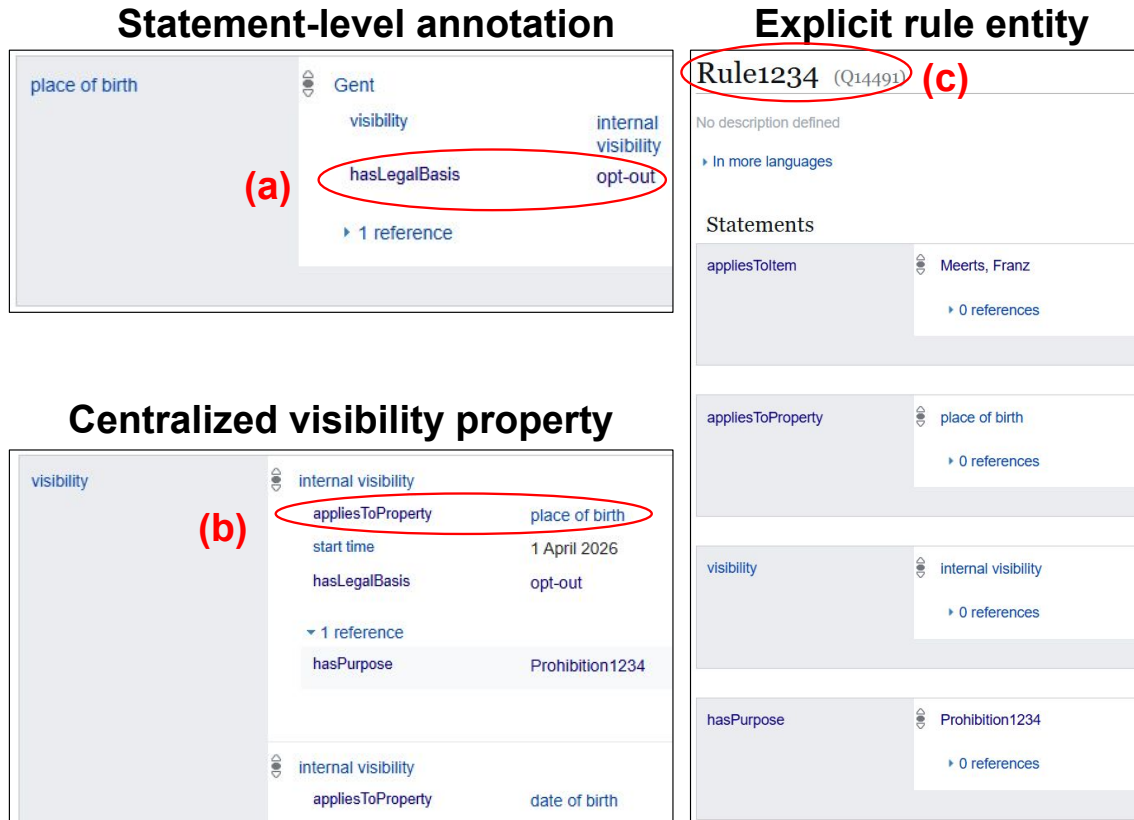


Figure 1: Three possible modelling approaches. (a) Whereas a statement-level annotation is straightforward, it has semantic issues as additional visibility qualifiers or temporal information qualify the property instead of the visibility. (b) A centralized visibility property is more expressive, but comes with modeling ambiguity. (c) Separate rule entities are very expressive but come with a high curation burden.

not even exist for the item. It is also ambiguous how to specify the visibility of the item itself, there are different options like simply omitting `appliesToProperty` or by using a property `appliesToItem`. But each comes with trade-offs as well.

Explicit rule entities. A third option separates governance into explicit rule entities that reference both the affected resource and the type of access decision (fig. 1, (c)). For instance, a rule item may state that *Rule123* applies to *Q23* and *P19*, sets *visibility* = *private*, and may include additional metadata such as a legal basis or validity interval. This approach is highly expressive and aligns with established standards such as ODRL, DPV, and PROV-O, since governance rules become first-class entities: the rule corresponds to an `odr1:Rule` (typically a Permission or Prohibition), the decision is expressed via an ODRL action scoped to a target, DPV provides structured legal grounding (e.g. `dpv:Consent`), and PROV-O can capture provenance of the rule (e.g. creation and attribution). However, it increases the number of entities that must be maintained, and is prone to error as it also introduces additional overhead for curators, i.e. one might create more than one rule entity with conflicting information and generally, the rule is not immediately visible on the item page.

4.1. Prototype implementation

The prototype presented in this work currently implements the first approach as a mediation layer in front of Wikibase. However, we currently investigate to implement all three approaches for greater flexibility. Data is retrieved via the Wikibase REST API and, if available, served from a Redis cache to improve response times. Visibility rules are applied at request time to ensure that only authorized content is returned. The properties used for *visibility* and *public* are configurable and not hard coded,

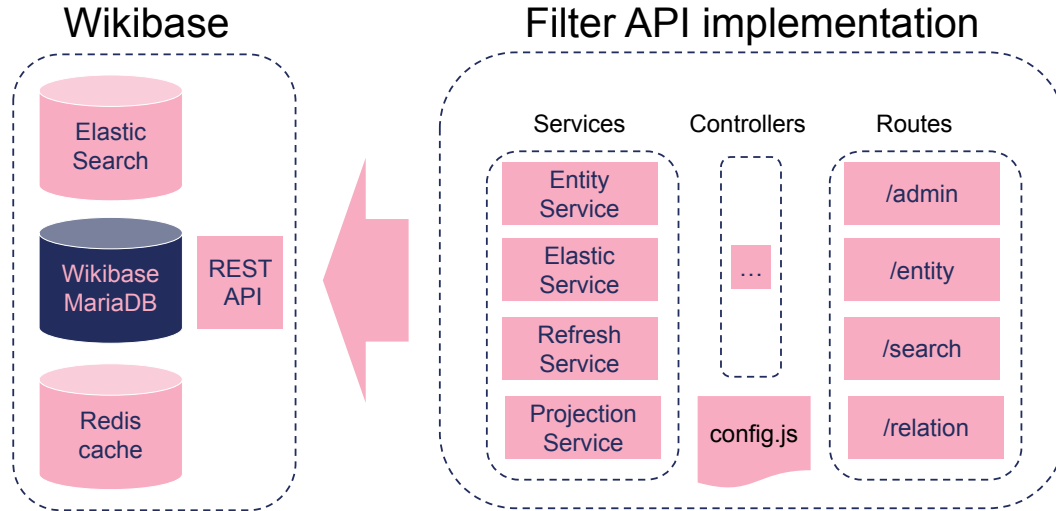


Figure 2: Architecture of the proposed system: entity data is retrieved from the Wikibase REST API and cached via Redis after applying filters at request time. ElasticSearch stores precomputed, governance-filtered property projections to support aggregated queries over public data.

allowing flexible use for different Wikibase instances.

In addition, ElasticSearch is used for selected property projections that support aggregated queries (e.g., retrieving all entities sharing a specific attribute such as place of birth). These projections are precomputed during indexing, where visibility rules are taken into account to ensure that only publicly accessible data is indexed. The standard Wikibase Query Service is not used, as its synchronization mechanism does not function well for our private Wikibase in the current MetaBelgica prototype deployment. The source code is available on GitHub⁸ and the main architecture is shown in fig. 2.

Figure 2 illustrates the overall system architecture. Requests for individual entities are resolved via the Wikibase REST API (or Redis cache), while ElasticSearch supports additional read-optimized projections for specific properties. The indexing configuration defines which properties are included in these projections, ensuring that aggregated retrieval scenarios remain efficient without exposing non-public data.

5. Conclusion and Future Work

This work is intended to support discussion on how fine-grained governance should be modelled in Wikibase-based knowledge graphs, rather than providing a final solution. The three approaches presented in the previous section highlight a spectrum between simplicity for curators, expressiveness of governance rules, and consistency of maintenance. While statement-level annotations are easy to apply locally, additional rule metadata is either not possible or if possible, not semantically sound. Centralised item-level visibility allows for more rule metadata, but introduces modelling ambiguity, particularly in cases where item-wide and property-level rules interact. Explicit rule entities provide the highest expressiveness, but shift complexity towards governance modelling and entity management.

A central observation is that increasing the expressiveness of governance modelling also increases the burden placed on curators. Statement-level annotations are intuitive and require little additional modelling effort, but provide limited support for complex governance requirements. At the other end of the spectrum, explicit rule entities can represent rich governance metadata, including legal basis and temporal validity, but introduce additional entities and relationships that curators must understand and maintain. The centralized visibility-property approach occupies a middle ground between these

⁸<https://github.com/MetaBelgica/wikibase-filter-api>

extremes. Determining which trade-off is most appropriate for collaborative cultural heritage curation remains an open question.

Rather than treating these approaches as isolated alternatives, we consider them as different trade-offs between curator effort, system complexity, and semantic clarity. We seek feedback on which of these trade-offs is most suitable in real-world cultural heritage settings.

Our implementation uses Elasticsearch only for precomputed, governance-filtered property projections that support aggregated queries, while direct entity access is served via the Wikibase REST API with Redis caching. This separation improves correctness for per-entity access and enables efficient cross-entity queries, but increases configuration overhead, since each indexed property must be explicitly defined within the config instead of the Wikibase.

Future work will focus on refining the proposed governance model based on community feedback and validating it within the current MetaBelgica environment. Particular attention will be given to the alignment of governance concepts with existing standards such as ODRL, DPV, and PROV-O, while preserving a modelling approach that remains understandable for non-technical curators. On the implementation side, further development of the prototype could explore tighter integration with the PHP-based Wikibase and MediaWiki stack to avoid the maintenance of two software stacks.

Declaration on Generative AI

During the preparation of this work, the author used GTP-5.5 in order to: **Grammar and spelling check** and **Improve writing style**. After using these services, the author reviewed and edited the content as needed and takes full responsibility for the publication's content.

References

- [1] S. Lieber, Opening up Belgian Cultural Heritage Reference Data with Wikibase, 2025. URL: <https://doi.org/10.5281/zenodo.17344191>. doi:10.5281/zenodo.17344191.
- [2] D. Diefenbach, M. D. Wilde, S. Alipio, Wikibase as an infrastructure for knowledge graphs: The eu knowledge graph, in: International Semantic Web Conference, Springer, 2021, pp. 631–647. URL: https://dl.acm.org/doi/10.1007/978-3-030-88361-4_37. doi:10.1007/978-3-030-88361-4_37.
- [3] L. Rossenova, P. Duchesne, I. Bluemel, Wikidata and Wikibase as complementary research data management services for cultural heritage data, in: Wikidata 2022: Wikidata Workshop 2022, Proceedings of the 3rd Wikidata Workshop 2022 co-located with the 21st International Semantic Web Conference (ISWC2022), 2022. URL: <https://doi.org/10.25968/opus-2573>. doi:10.25968/opus-2573.
- [4] H. J. Pandit, B. Esteves, G. P. Krog, P. Ryan, D. Golpayegani, J. Flake, Data privacy vocabulary (dpv)–version 2.0, in: International Semantic Web Conference, Springer, 2024, pp. 171–193. URL: https://dl.acm.org/doi/10.1007/978-3-031-77847-6_10. doi:10.1007/978-3-031-77847-6_10.
- [5] T. Lebo, S. Sahoo, D. McGuinness, K. Belhajjame, J. Cheney, D. Corsar, D. Garijo, S. Soiland-Reyes, S. Zednik, J. Zhao, PROV-O: The PROV Ontology, Recommendation, World Wide Web Consortium (W3C), 2013. URL: <https://www.w3.org/TR/prov-o/>.
- [6] R. Iannella, S. Villata, ODRL Information Model 2.2, W3C Recommendation, W3C, 2018. URL: <https://www.w3.org/TR/2018/REC-odrl-model-20180215/>.
- [7] D. Antal, Wikibase as a data sharing space: Connecting rights, communities, and glam through federated infrastructures, 2025. URL: <https://doi.org/10.5281/zenodo.17496740>. doi:10.5281/zenodo.17496740.