

Semantic MediaWiki as a Governed Architecture for Ontology-Driven Collaborative Knowledge Engineering

Rajesh Piryani^{1,*}, Yannick Hersant¹, Benoit Dalet¹, Carole Convers¹, Benoit Baurens^{1,*} and Abdessadeq Zougari Ben Elkhayat¹

¹AKKODIS Research, AKKODIS, Blagnac 31700, France

Abstract

Semantic technologies provide powerful mechanisms for structured knowledge representation, interoperability, and machine interpretability. However, their integration into collaborative and operational knowledge management environments remains limited due to usability, governance, and validation challenges. This paper presents the SCIMIN CRM Open Collaborative Knowledge Base, a Semantic MediaWiki (SMW) based platform extended with ontology driven semantic annotation, AI assisted enrichment, validation workflows, and controlled publication mechanisms. The system combines wiki-based collaborative authoring with formal semantic modeling grounded in a domain ontology and RDF representations. We describe the system's logical, physical, and semantic architecture, its ontology-aligned knowledge model, and its hybrid human-machine annotation workflow. The SMW platform applied as an open collaborative semantic knowledge base infrastructure supporting structured knowledge acquisition, semantic annotations and querying, RDF export, and governance across integration and production environments. The work contributes architectural design insights and practical lessons for deploying ontology based semantic systems in collaborative environment.

Keywords

Semantic MediaWiki, knowledge graphs, ontology engineering, semantic annotation, collaborative knowledge base, RDF, SPARQL, human-in-the-loop AI

1. Introduction

Semantic technologies enable structured knowledge representation, interoperability, and machine interpretability across distributed information systems. Semantic wikis extend traditional wiki platforms by embedding machine-readable annotations within collaboratively authored content, thereby supporting semantic navigation, structured search, and knowledge reuse [1]. Platforms such as Semantic MediaWiki (SMW) [2, 3, 4], IkeWiki [5, 6], and SemperWiki [7] have demonstrated the potential of combining collaborative authoring with Semantic Web technologies, including RDF [8] and OWL [9]. More recently, domain-specific deployments have applied SMW to scientific and educational knowledge management scenarios [10, 11]. In parallel, large language models (LLMs) have enabled the automated extraction of entities and relationships from unstructured text, while ontology-driven approaches have emphasized the importance of ontology grounding and validation to ensure semantic consistency and reliability [12]. Despite these advances, challenges remain in ensuring semantic consistency, validation, governance, and interoperability within collaborative semantic environments. In particular, the integration of semantic wiki platforms, AI-assisted semantic annotation, human validation, and governed publication workflows remains underexplored.

The SCIMIN CRM (Sustainable and Circular Production of Mineral Critical Raw Materials) project, funded under the Horizon Europe Research and Innovation Programme provides a representative multi-stakeholder setting in which these challenges arise. The project seeks to enhance circularity and sustainability within the aggregates and mining sector by characterizing, valorizing, and recov-

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

*Corresponding author.

✉ rajesh.pirayani@akkodis.com (R. Piryani); Yannick.HERSANT@akkodis.com (Y. Hersant); benoit.dalet@akkodis.com (B. Dalet); Carole.CONVERS@akkodis.com (C. Convers); benoit.baurens@akkodis.com (B. Baurens); sadeq.zougari@akkodis.com (A. Zougari Ben Elkhayat)

ORCID: 0000-0003-3374-0657 (R. Piryani); 0000-0002-4823-205X (B. Baurens); 0009-0003-7369-6555 (A. Zougari Ben Elkhayat)



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

ering historical mining stockpiles that were previously underutilized. Through the development of novel tools, assessment methodologies, and recovery processes, SCIMIN CRM contributes to reducing dependence on external raw materials while supporting environmental rehabilitation and resource efficiency. The project generates a substantial corpus of heterogeneous knowledge artefacts including technical documents, environmental assessments, sustainability analyses, data services, and guidelines which require collaborative management, structured categorization, interoperable publication, and controlled validation. To address these needs, the SCIMIN-CRM digital platform adopts a Semantic MediaWiki-based approach supported by a domain ontology [13, 14], semantic search capabilities, and interoperability through standardized APIs. Within this framework, SMW serves as the foundation for combining collaborative authoring with ontology-grounded semantic annotation and Linked Data exposure.

This paper presents the SCIMIN CRM Open Collaborative Knowledge Base, an extended SMW platform that integrates ontology-driven semantic annotation, AI-assisted enrichment, validation workflows, and controlled synchronization between integration and production environments. Its main contributions are: (i) a service-oriented semantic architecture combining ontology alignment and AI-based enrichment within a Wikimedia-style CMS; (ii) a hybrid human-machine validation workflow supporting semantic annotation validation and controlled publication; and (iii) practical insights into deploying an ontology-based knowledge base in a Horizon Europe research context. The platform demonstrates how SMW infrastructures can enable structured, interoperable, and sustainable knowledge management for European research initiatives.

2. Methodology

2.1. Architecture

The proposed platform extends SMW with services for document ingestion, ontology-driven semantic annotation, validation, and controlled publication. As illustrated in Figure 1, the system establishes an end-to-end knowledge engineering workflow that transforms unstructured documents into ontology-aligned and machine-interpretable knowledge. In addition to collaborative content authoring, the platform supports semantic querying, RDF generation, and interoperability through standardized APIs. To ensure traceability and governance, the system is deployed across separate integration and production environments (Figure 2). Content creation, semantic annotation, and validation are primarily conducted in the integration environment, while validated content can be synchronized to the production environment through a controlled publication workflow. These operations may also be performed directly in the production environment when required.

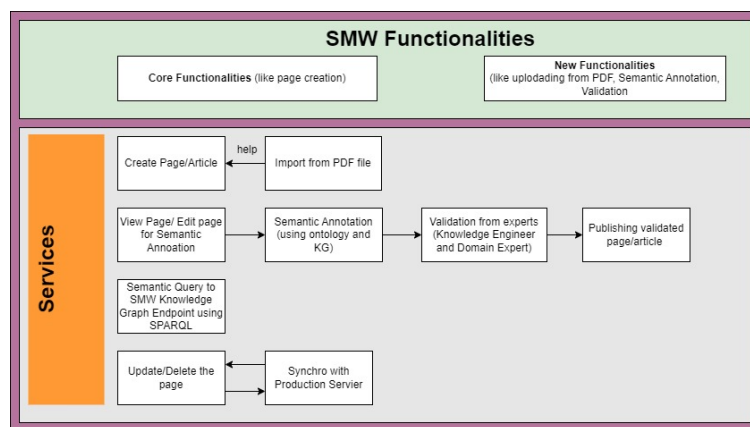


Figure 1: Architecture block diagram of SMW.

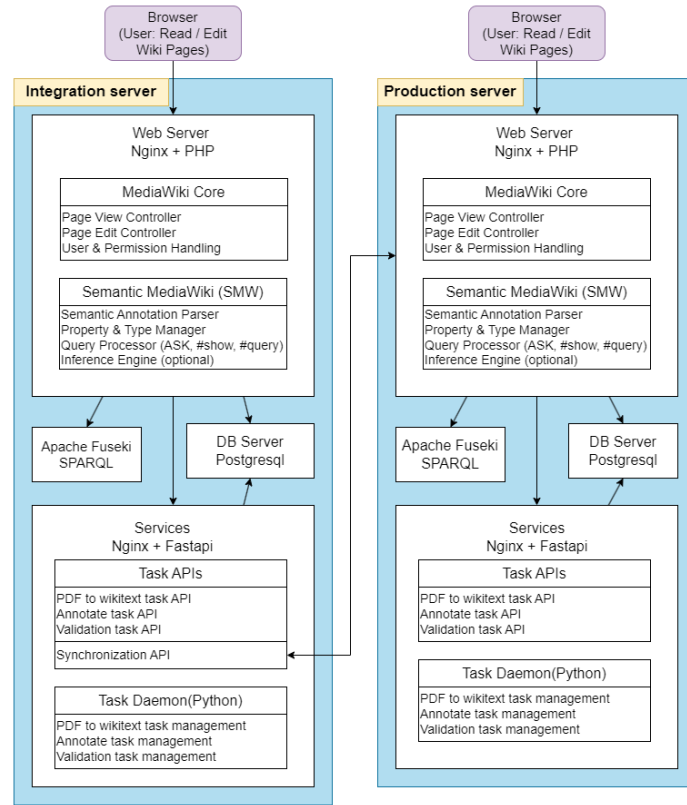


Figure 2: Physical/Deployment architecture of SMW.

2.2. Ontology driven semantic annotation

The semantic annotation process is grounded in the SCIMIN CRM ontology [13, 14], which defines the classes, object properties, and data properties used to structure domain knowledge. In SMW categories are aligned with ontology classes, relations with object properties, and attributes with datatype properties [1]. Table 1 summarizes the correspondence between SMW annotations and ontology elements.

Table 1
Mapping between Semantic MediaWiki and ontology elements

SMW Element	Ontology Interpretation	Example
Category	Classes	[[Category:MiningWasteFacility]]
Property of type page	Object properties	[[located at::Weissenbach]]
Property of type text/number/boolean	Data properties	[[has Area::approximately 6 KM ²]]

Source documents are first transformed into WikiText and processed using an LLM. As shown in Figure 3, ontology concepts are incorporated into automatically generated prompts to guide concept and relation extraction. The generated annotations are aligned with ontology elements, validated through ontology-conformance, hallucination metrics [12] and expert review, and finally encoded as SMW annotations that can be exported as RDF and queried through SPARQL.

2.3. Validation and governance workflow

For validation, generated annotations undergo a two-stage validation process. First automated validation verifies ontology, including domain and range constraints, datatype consistency, and structural correctness of generated triples and hallucination metrics[12]. Additional checks are performed to identify

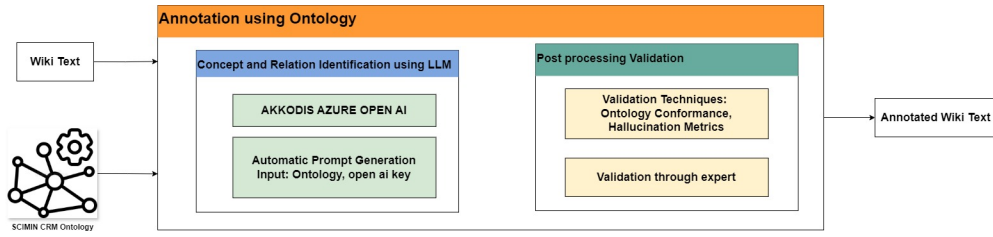


Figure 3: Semantic Annotation of Wiki text using Ontology.

LLM-generated annotations that are inconsistent with either the source text or ontology definitions [12]. Next, domain experts review candidate annotations through a human-in-the-loop (HITL) workflow, approving, modifying, or rejecting them before publication. Only annotations that successfully pass both automated and expert validation are synchronized to the production environment. By combining ontology grounding, AI-assisted annotation, and expert validation, the proposed workflow supports reliable and governed collaborative knowledge engineering while maintaining semantic consistency and interoperability.

3. Results

The proposed platform was deployed within the SCIMIN CRM project to support collaborative knowledge acquisition, semantic enrichment, and governed publication of domain knowledge. The demonstration highlights the complete workflow from document ingestion to ontology-aligned knowledge publication. The platform is implemented using SMW, PostgreSQL, Apache Fuseki, FastAPI-based annotation services, and Azure OpenAI models deployed on Azure infrastructure.

3.1. Document ingestion

The platform incorporates an AI-assisted ingestion service that transforms PDF documents into editable WikiText pages within SMW. Using the PyMuPDF4LLM library and LLM-based parsing, document content is automatically converted while preserving its structure and embedded elements. Users can review and refine the generated content before publication, reducing manual transcription effort and facilitating subsequent semantic enrichment. Figure 4 illustrates the creation of wiki content from a PDF document.

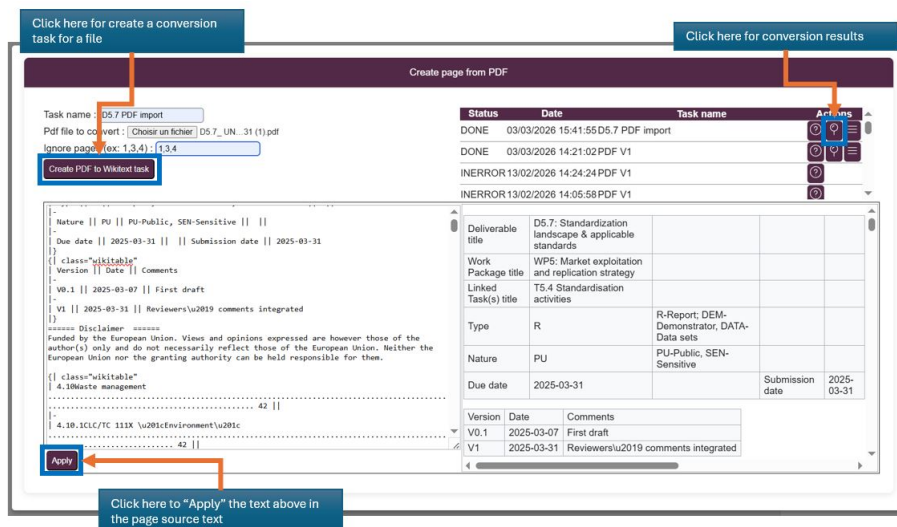


Figure 4: PDF-to-WikiText content ingestion workflow.

Category: MiningWasteFacility	
Facts	
... more about "PILOT 1- South (Regional case)" RDF feed	
Scimin:hasCharacteristics	Geotechnical properties +
Scimin:hasLocation	Llanera + , Asturias + and Spain +
Scimin:hasMW	Tailings + and WasteRock +
Scimin:hasMWMaterial	Fluorspar + , Baryte + , Silica + and Calcite +
Scimin:hasStatus	Abandoned +
Scimin:hasVolume	2000000 tonnes +

Figure 5: Example of ontology-aligned semantic annotations represented as structured facts.

3.2. Ontology driven semantic annotation and validation

The platform supports ontology-driven semantic annotation through an LLM-assisted workflow integrated into SMW. Users can generate semantic annotations for selected content sections, while ontology concepts from the SCIMIN CRM ontology[13, 14] guide entity and relation extraction. Generated annotations are reviewed through a HITL validation process that separates annotation generation from approval. Approved annotations are incorporated into wiki pages as structured semantic facts linked to ontology concepts such as mining waste facilities, materials, activities, and locations. Figure 5 presents an example of ontology-aligned semantic annotations represented as structured facts within SMW. These semantic assertions can be exported as RDF and queried through SPARQL, enabling knowledge graph generation and interoperability.

3.3. Governance and Controlled Publication

To ensure data quality and traceability, the platform implements a governance-oriented publication workflow. Content creation, annotation, and validation are performed in the integration environment, while only validated content is propagated to the production environment through a controlled synchronization process. The synchronization mechanism provides revision tracking, version comparison, and approval management, ensuring that only verified knowledge becomes publicly accessible. This separation of environments supports traceability, iterative refinement, and reliable publication of semantic content. Overall, the results demonstrate the feasibility of combining SMW, ontology-driven modeling, LLM-assisted annotation, and human validation within a governed collaborative knowledge engineering workflow.

4. Conclusion

This research work presents the SCIMIN CRM Open Collaborative Knowledge Base, a governance-oriented Semantic MediaWiki platform integrating ontology-driven knowledge modeling, AI-assisted document ingestion, and semantic annotation within a controlled workflow that supports validation and publication across distributed environments. The proposed approach transforms unstructured documents into ontology-aligned, machine-interpretable knowledge through ontology grounding and HITL validation. The results demonstrate the feasibility of combining semantic web technologies, large language models, and collaborative authoring while ensuring consistency, traceability, and interoperability. Future work will focus on quantitative evaluation, scalability assessment, and the integration of advanced validation and reasoning mechanisms. SCIMIN CRM project knowledge Base aims to become a reference system aligned with the anticipated publication of the first MWF standards by 2030, ensuring lasting impact beyond the funded period.

Acknowledgments

This study has been performed under the project SCIMIN-CRM-Sustainable & Circular Production of Mineral Critical Raw Materials, funded by Horizon Europe research and innovation programme (UE:101177746).

Declaration on Generative AI

In this research work, the authors used generative AI tools (Copilot) during the preparation of the manuscript for grammar and spelling correction, as well as for paraphrasing and rewording. The authors have carefully reviewed and edited the generated content where necessary and take full responsibility for the final content of the publication.

References

- [1] M. Kroetzsch, S. Schaffert, D. Vrandecic, Reasoning in semantic wikis, in: Reasoning Web International Summer School, Springer, Berlin, Heidelberg, 2007, pp. 310–329.
- [2] M. Volkel, S. Schaffert (Eds.), 1st Workshop From Wiki to Semantics (SemWiki'06), Budva, Montenegro, 2006.
- [3] M. Kroetzsch, D. Vrandecic, M. Volkel, Semantic mediawiki, in: Proceedings of the 5th International Semantic Web Conference, 2006, pp. 935–942.
- [4] M. Kroetzsch, D. Vrandecic, M. Volkel, Wikipedia and the semantic web – the missing links, in: Proceedings of WikiMania, 2005.
- [5] S. Schaffert, Ikewiki: A semantic wiki for collaborative knowledge management, in: STICA 2006, Manchester, UK, 2006.
- [6] S. Schaffert, R. Westenthaler, A. Gruber, Ikewiki: A user-friendly semantic wiki, 2006. Demonstration at ESWC 2006.
- [7] E. Oren, Semperwiki: A semantic personal wiki, in: Workshop on the Semantic Desktop, Galway, Ireland, 2005.
- [8] D. Brickley, R. V. Guha, Rdf vocabulary description language 1.0: Rdf schema, 2004. W3C Recommendation.
- [9] D. McGuinness, F. van Harmelen, Owl web ontology language overview, 2004. W3C Recommendation.
- [10] C. Neidiger, T. Saier, K. Kuehn, V. Larignon, M. Faerber, C. Bizzarri, S. Braese, Implementation of an open chemistry knowledge base with a semantic wiki, *Journal of Cheminformatics* 17 (2025) 99.
- [11] O. Pinchuk, J. Rogushina, L. Kondratova, Developing digital competence of the ukrainian electronic encyclopedia of education users based on semantic wiki technologies, in: CEUR Workshop Proceedings, volume 4096, 2025, pp. 35–52.
- [12] N. Mihindukulasooriya, S. Tiwari, C. F. Enguix, K. Lata, Text2kgbench: A benchmark for ontology-driven knowledge graph generation from text, in: International Semantic Web Conference, Springer, Cham, 2023, pp. 247–265.
- [13] SCIMIN CRM Ontology, 2026. Available online: <https://sadeqzougari.github.io/scimin-crm/>. Accessed: May 2026.
- [14] R. Piryani, X. Xiao, K. Bhadani, G. Asbjornsson, M. Evertsson, A. Zougari Ben Elkhayat, AI-driven development of mining waste facilities ontology, 2026. Communicated in IPMC 2026 Conference.