Assessing Knowledge Provenance Quality for CEO Decision Support: An OWL-DL Ontological Framework

Carolina de Oliveira and Fernando Alvaro Ostuni Gauthier
Santa Catarina Federal University, Florianópolis, Brazil

carolina.oliveira.c.o@posgrad.ufsc.br
fernando.gauthier@ufsc.br

Abstract: The ontology demonstrates approaches to enhancing the precision of knowledge provenance and outlines the systematic engineering of a knowledge provenance ontology designed to improve its reliability for executive decision-making. Utilizing architectural principles, we propose a formal model that evaluates knowledge provenance quality based on standardized coding systems such as PROV-O. The methodology follows a rapid prototyping lifecycle, employing the "Ontology 101" development process to transform domain expertise into a rigorous OWL-DL framework. By implementing constraints and functional properties, the ontology ensures data integrity and consistency, which are critical for CEOs navigating complex information environments. The results demonstrate that formal ontological reasoning, supported by tools such as Pellet, can effectively filter low-quality knowledge sources, providing a trustworthy basis for strategic corporate oversight.

Keywords: PROV-O, Ontology, Quality, Knowledge engineering

1. Problem

Organizational decision-making at all levels relies on tacit and explicit knowledge, which must be high-quality and verifiable, especially when involving Agentic AI (Wang and Leem, 2026). While universal data quality dimensions include completeness, correctness, consistency, and timeliness (Reda et al., 2023; Secundo et al., 2024; Wang and Leem, 2026), organizational knowledge quality (KQ) requires distinct dimensions, including adaptability, applicability, expandability, truth, innovativeness, and justification (Waheed and Kaur, 2016).

Furthermore, knowledge provenance quality enables informed decision-making by providing transparent evidence of changes. Documenting an item's origin serves as a critical trust indicator (Choi and Cheng, 2025). This process-centered provenance captures the specific actions and production steps taken to generate information (Gil and Miles, 2013; Herschel et al., 2017). This transparency is vital for Explainable AI (Kale et al., 2023) and high-stakes decision-making (Sode, 2026).

The objective of this study is to develop an ontology to represent the characteristics of knowledge provenance quality, addressing the central question: How can we assess the knowledge provenance quality used in business processes?

2. Methodology

The decision to use a heavyweight ontology was made with the aim of building a shareable and reusable knowledge base and representing the “processes and activities for solving a given problem by abstracting the domain context” (Isotani and Bittencourt, 2015).

This study adopted OWL-DL, a W3C standard sublanguage, for ontology modeling because of its maximum conceptual expressiveness, computational completeness, and decidability. Positioned between propositional and first-order logic, OWL-DL provides key capabilities, such as automatic inference, cardinality constraints, and semantic interoperability, which motivated its selection.

The implementation tool selected was Protégé (Musen, 2015), version 5.6.7, because it is a collaborative OWL ontology editor. The development of the ontology followed the seven steps of the “Ontology 101” process (Noy and McGuinness, 2001), described below.

1. Domain and scope: The ontology’s domain is knowledge provenance, and its scope is to determine whether a source is “good” or “bad,” taking into account the following competency questions:

- 1.1 Who has the authority to create knowledge?
- 1.2 Which department within the organisation is responsible for creating knowledge?
- 1.3 When was the knowledge created?
- 1.4 What is the knowledge provenance: human [tacit], machine, or media [explicit]?
- 1.5 What is the quality of the knowledge provenance?

2. Reuse of existing resources: The W3C's PROV-O standard was used to ensure that this ontology is interoperable with international standards (Lebo *et al.,* 2013).
3. List of terms: Table 1.

Table 1: Concepts dictionary (draft)

Concept	PROV-O (W3C)	The Authors
Entity, Activity, Agent, Influence, InstantaneousEvent, Location, Role	https://www.w3.org/TR/prov-o/#cross-reference-index	--
KnowledgeProvenance	--	Knowledge about its production process.
ProvenanceQuality	--	Characteristics that address the essential requirements for understanding its production process.

4. Class hierarchy using a top-down approach: (Figure 1).

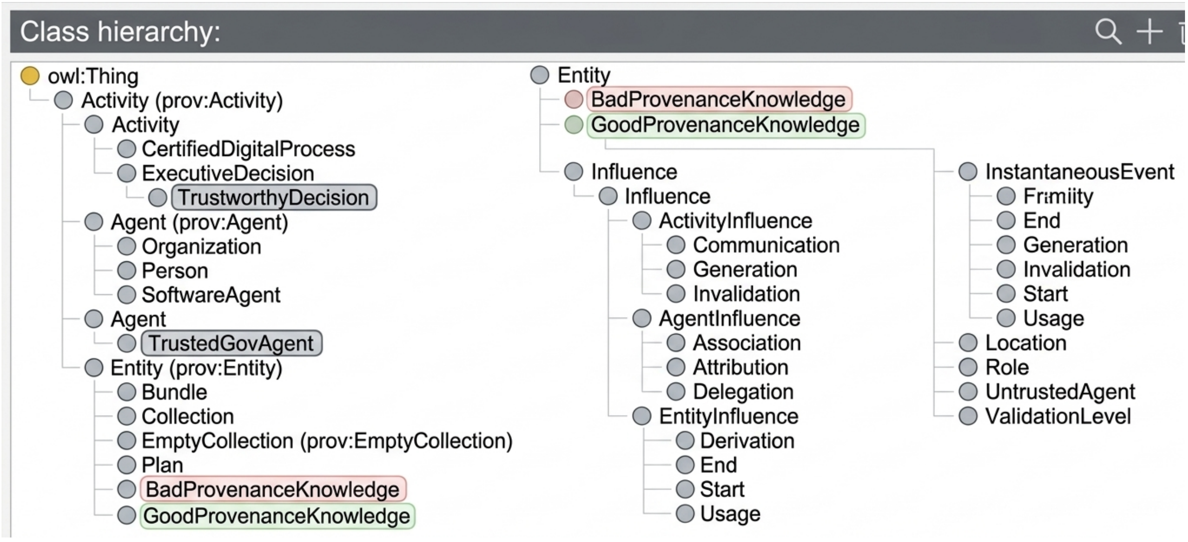

Figure 1: The world of Knowledge Provenance (Protégé).

5. Definition of properties: Object properties were imported from the W3C's PROV-O standard. To satisfy the integrity and consistency requirements, functional properties such as "hasQualityScore" were defined. In OWL, this means an individual can have at most one quality score, preventing contradictory "Good" and "Bad" labels for the same entity.
6. Facets definition: In the decision-making process, a "poor" source of knowledge results from contradictory, missing, or fraudulent information.
7. Instance creation: The process of populating the assertional box (A-Box), moving from an abstract model to the specific real-world model that the ontology represents.

The ontology categorizes knowledge provenance quality as "Good" or "Bad" based on the intersection of three dimensions: (1) it must be generated by a "CertifiedDigitalProcess"; (2) it must be attributed to a "TrustedGovAgent"; and (3) it must not have any links to an "UntrustedAgent." Through formal reasoning, an entity is classified as "Good" only if it satisfies the following Description Logic criteria:

$$GoodProvenanceKnowledge \equiv Entity \sqcap \exists\, wasGeneratedBy.CertifiedProcess \sqcap \forall\, wasAttributedTo.TrustedAgent \sqcap \exists\, hasQuality.\{Timely, Complete, Certified\}$$

To establish a functional binary classification gate within an arbitrary coding exercise, these three criteria were pragmatically identified to directly exploit the structural primitives of the W3C PROV-O standard, specifically by mapping them to its core Activity, Agent, and Entity components. The first criterion, requiring generation by a "CertifiedDigitalProcess," isolates the data's procedural history, enabling the script to algorithmically distinguish audited automated pipelines from unverified manual entries.

The second criterion, attribution to a "TrustedGovAgent," establishes a strict origin checkpoint by verifying structural authority through the standard "prov:wasAttributedTo" object property. The final criterion, requiring the complete absence of links to an "UntrustedAgent," was deliberately introduced to operationalize universal quantification ("owl:allValuesFrom") within an OWL-DL environment. This design choice functions as a strict logical contamination filter; if an entity is linked to an unauthorized source, it triggers an immediate logical failure.

Taken together, these three dimensions form a rigid Boolean intersection within Description Logic (DL). This framework enables an OWL reasoner such as Pellet to automatically and deterministically determine whether incoming data instances belong to the "Good" or "Bad" provenance classes, thereby providing a clean, algorithmic gate for an executive decision-making dashboard without requiring manual oversight or subjective human evaluation.

3. Results

The metrics reported by Protégé are as follows: 42 classes, 6 individuals, 48 object properties, and 6 data properties. The class axioms comprise 32 subclasses, 3 equivalent classes, and 5 disjoint classes.

The main goal of the ontology is to explain how to assess the knowledge provenance quality used in business processes. We present an illustrative use case involving the formulation of public policies aimed at regulating the use of artificial intelligence agents within public servants' workflows. The ontology indicates whether the knowledge possesses the qualities necessary for decision-makers to proceed with or halt a course of action.

This OWL-DL framework provides a real-time, automated trust filter for a CEO leading digital transformation, for example. Instead of manually auditing complex data lineages, the executive relies on the reasoner to automatically flag information quality issues. By dynamically restricting strategic decisions to only consume certified, complete, and timely knowledge, the ontology creates an explicit mathematical safety gate.

This ensures that high-stakes public sector implementations are protected from unverified sources and outdated data, directly mitigating critical strategic governance risks.

4. Discussion and Contribution

Evaluating information is challenging due to contradictory content, making knowledge provenance vital for trust (Moreau and Missier, 2013). While recent studies have explored ontology quality assessment (Wilson *et al.*, 2023) and the tracking and assessment of provenance, knowledge quality remains a persistent real-world challenge (Huynh *et al.*, 2021; Osman, Mohd Noah and Saad, 2022; Zhang, Zhou and Ujcich, 2024; Prudhomme *et al.*, 2025; Arora *et al.*, 2025)

In high-stakes corporate or government scenarios, Knowledge Engineering addresses this challenge by converting raw data into "actionable knowledge" through structured ontologies. These ontologies support explainable Hybrid Intelligence, reducing human cognitive biases (Tiddi *et al.*, 2023; Said *et al.*, 2025). When combined with Trust Recommender Systems—which use social trust networks—this approach ensures that recommendations are transparent, highly credible, and aligned with trusted expert consensus, ultimately aiding executives such as CEOs in confident decision-making (Akdim, 2026). To ensure system integrity, automated reasoners (such as Pellet) evaluate knowledge provenance quality so both human and AI agents work solely with consistent and valid data.

This study introduces an ontological framework designed to assess knowledge provenance quality and mitigate misinformation risks. The main limitation of this study is the lack of real-world application.

5. Conclusion and Future Works

This study is a work in progress within a thesis. Assessing knowledge provenance quality using OWL-DL provides a rigorous and scalable solution, and future studies will refine the framework by implementing it within an eligible Brazilian Federal Executive Branch unit while considering the role of ontology and hybrid-intelligent environments (Pileggi, 2024).

Ethics Declaration: No ethical clearance was required for this research.

AI Declaration: An AI tool was used for summarizing some parts of the paper, and creating figures, with careful decisions made by the authors.

References

Akdim, I., Mekouar, L. and Iraqi, Y. (2026) 'Trust in recommender systems: A survey.' *Expert Systems with Applications*, 298, 129653. https://doi.org/10.1016/j.eswa.2025.129653.

Arora, M. *et al.* (2025) 'Case study: Raising the bar on content quality at Amazon Web Services.' https://www.apqc.org/resource-library/resource/raising-bar-content-quality-amazon-web-services/html.

Choi, I. and Cheng, Y. Y. (2025) 'A conceptual model for tracking the provenance of activities in knowledge organization systems.' *Journal of Documentation*, 81(1), 147–167. https://doi.org/10.1108/JD-05-2024-0101.

Gil, Y. and Miles, S. (2013) 'PROV Model Primer.' W3C. http://www.w3.org/TR/2013/NOTE-prov-primer-20130430/.

Herschel, M., Diestelkämper, R. and Ben Lahmar, H. (2017) 'A survey on provenance: What for? What form? What from?' *The VLDB Journal*, 26(6), 881–906.

Huynh, T. D. *et al.* (2021) 'Addressing regulatory requirements on explanations for automated decisions with provenance—A case study.' *Digital Government: Research and Practice*, 2(2), 1–14. https://doi.org/10.1145/3436897.

Isotani, S. and Bittencourt, I. I. (2015) *Dados abertos conectados*. Novatec Editora. http://ceweb.br/publicacao/livro-dados-abertos/.

Kale, A. *et al.* (2023) 'Provenance documentation to enable explainable and trustworthy AI: A literature review.' *Data Intelligence*, 5(1), 139–162. https://doi.org/10.1162/dint_a_00119.

Lebo, T., Sahoo, S. and McGuinness, D. (2013) 'PROV-O: The PROV Ontology.' https://www.w3.org/TR/2013/REC-prov-o-20130430/.

Moreau, L. and Missier, P. (2013) 'PROV-DM: The PROV Data Model.' https://www.w3.org/TR/2013/REC-prov-dm-20130430/#dfn-provenance.

Musen, M. A. (2015) 'The Protégé project: A look back and a look forward. AI Matters.' *Association of Computing Machinery Specific Interest Group in Artificial Intelligence*, 1(4).

Noy, N. F. and McGuinness, D. L. (2001) *Ontology development 101: A guide to creating your first ontology*. Stanford Knowledge Systems Laboratory Technical Report KSL-01-05.

Osman, M. A., Mohd Noah, S. A. and Saad, S. (2022) 'Ontology-based knowledge management tools for knowledge sharing in organization—A review.' *IEEE Access*, 10, 43267–43283. https://doi.org/10.1109/ACCESS.2022.3163758.

Pileggi, S. F. (2024) 'Ontology in hybrid intelligence: A concise literature review.' *Future Internet*, 16(8), 268. https://doi.org/10.3390/fi16080268.

Prudhomme, T. *et al.* (2025) 'A semantic approach to mapping the Provenance Ontology to Basic Formal Ontology.' *Scientific Data*, 12(1), 282. https://doi.org/10.1038/s41597-025-04580-1.

Reda, O., Benabdellah, N. C. and Zellou, A. (2023) 'A systematic literature review on data quality assessment.' *Bulletin of Electrical Engineering and Informatics*, 12(6). https://doi.org/10.11591/eei.v12i6.5667.

Secundo, D. *et al.* (2024) *Study of the relationships between the dimensions of information quality and knowledge management processes in forensic processes at crimes against life locations*. https://doi.org/10.4236/jss.2024.126005.

Said, M. E. *et al.* (2025) 'Ontology-based knowledge integration framework: A systematic review.' *International Journal on Advanced Science, Engineering and Information Technology*, 15(6), 1973–1979. https://doi.org/10.18517/ijaseit.15.6.20819.

Sode, K. (2026) 'The interpreter's trap: How explainable AI launders uncertainty into justification—a socio-legal case study of COMPAS risk assessment.' *AI and SOCIETY*, 41(5), 5285–5299. https://doi.org/10.1007/s00146-026-02885-2.

Tiddi, I. *et* al. (2023) 'Knowledge engineering for hybrid intelligence.' *Proceedings of the 12th Knowledge Capture Conference 2023*, 75–82. https://doi.org/10.1145/3587259.3627541.

Waheed, M. and Kaur, K. (2016) 'Knowledge quality: A review and a revised conceptual model.' *Information Development*, 32(3), 271–284. https://doi.org/10.1177/0266666914539694.

Wang, N. and Leem, S. -G. (2026) 'Modernizing data quality: Evolving dimensions for unstructured text data in big data, AI and ethical contexts.' *Data Quality Matters – Best Practices for Integrity and Assurance.* IntechOpen. https://doi.org/10.5772/intechopen.1013232.

Wilson, R. S. I. *et al.* (2023) 'A conceptual model for ontology quality assessment: A systematic review.' *Semantic Web*, 14(6), 1051–1097. https://doi.org/10.3233/SW-233393.

Zhang, J., Zhou, W. and Ujcich, B. E. (2024) 'Provenance-enabled explainable AI', *Proceedings of the ACM on Management of Data*, 2(6), 1–27. https://doi.org/10.1145/3698826.