MMO: A Lightweight Semantic and Trust Model for Metadata

Valentin Mouche
Université de Pau et des Pays de l'Adour, E2S UPPA, LIUPPA
Mont de Marsan, France

Nouha Laamech
University Mohammed VI Polytechnic, COLCOM
Rabat, Morocco
Nouha.laamech@um6p.ma

Manuel Munier
Université de Pau et des Pays de l'Adour, E2S UPPA, LIUPPA
Mont de Marsan, France
Manuel.munier@univ-pau.fr

Abstract

Metadata play a critical role in structuring and contextualizing data, particularly in security systems where provenance, quality, and traceability directly influence the reliability and accountability of downstream models. However, definitions and practices around metadata remain fragmented across domains, limiting their reuse and interoperability. This paper presents a lightweight, extensible ontology for metadata, based on Semantic Web standards, metadata are designed to support secure data exchange, enhance trust through semantic transparency, and enable robust provenance tracking across complex information systems. Covering key metadata dimensions, descriptive, structural, administrative, and temporal, the ontology also incorporates provenance and dynamic metadata, enabling auditable data flows and consistent handling to strengthen data integrity, traceability, and security in critical systems.

CCS Concepts

• **Security and privacy → Formal security models**; **Formal methods and theory of security**.

Keywords

Metadata, Semantic Model, Security and Trust, Knowledge Management

ACM Reference Format:
Valentin Mouche, Nouha Laamech, and Manuel Munier. 2025. MMO: A Lightweight Semantic and Trust Model for Metadata . In *The 6th Workshop on Secure IoT, Edge and Cloud systems (SIoTEC '25), November 10–14, 2025, Seoul, Republic of Korea.* ACM, New York, NY, USA, 6 pages. https://doi.org/10.1145/3770740.3771879

1 Introduction

Data-driven applications increasingly depend on the collaboration of multiple stakeholders to achieve their objectives efficiently and reliably. These applications typically involve the integration of heterogeneous information systems that must interoperate seamlessly and share data across organizational or technical boundaries. For instance, in the energy domain, various entities such as smart buildings, industrial facilities, and energy providers exchange consumption data in order to optimize resource allocation, reduce waste, and promote the integration of renewable energy sources. In such distributed environments, the reliability and trustworthiness of shared data become a critical concern, particularly for applications in regulated or security-sensitive sectors [1, 21]. Traditional metrics used to evaluate data quality, namely confidentiality, integrity, availability, and traceability, are necessary but insufficient [14]. These properties do not fully capture the semantic, contextual, or operational characteristics of data that influence its reliability in practice.

To address these limitations, recent research emphasizes the role of metadata for security as a first-class component in assessing data quality [6]. According to the National Information Standards Organization (NISO), metadata are defined as "structured information that describes, explains, locates, or otherwise represents something else" [22]. This definition emphasizes the diverse roles metadata can play, ranging from descriptive and explanatory to locational and representative functions. Crucially, metadata are not static annotations: they are inherently contextual and dynamic, shaped by the nature of the data they describe and the environment in which they are used. For instance, it can be used to enhance decision making [8], the transparency of open data [23], or the basement of machine-learning security algorithms [11].

Metadata do more than only describing data, it enables easier data discovery, and can help increase understanding of a given information. For example, in a smart city, air quality data collected by sensors is accompanied by metadata such as location, timestamp, and sensor status. Additionally, metadata can offer contextual clues that help users better grasp the content and relevance of the data they are examining, ensuring accurate interpretation across information systems [17]. We advocate that detailed, semantically rich metadata can be considerered as security and trust metrics. Beyond describing content or format, metadata must capture operational and contextual information such as provenance, access control, usage constraints, and applied transformations. For instance, being able to trace the lineage of data from raw system logs to processed indicators is critical for validating incident reports, conducting forensic analysis, or ensuring compliance with security policies. Metadata can also support early detection of anomalies, inconsistencies, or policy violations, before they propagate through systems [9]. Although the importance of metadata quality has gained increasing attention in recent years, there is still no universally accepted formalism. Definitions of metadata quality are often highly domain-specific and tend to vary significantly across different domains. Moreover, this lack of standardization challenge interoperability

SIoTEC '25, Seoul, Republic of Korea

ACM ISBN 978-1-4503-XXXX-X/2018/06
https://doi.org/10.1145/3770740.3771879

and makes it difficult to assess or compare metadata quality across heterogeneous and distributed systems.

The establishment of a formal knowledge representation of metadata can play a critical role in addressing this challenge by enabling context-aware evaluation of data and ensure semantic interoperability [29]. Through appropriate metadata descriptions, decision makers can better judge whether the data at hand are "fit for purpose" in the specific context of their task. Situating data quality within a semantically enriched and formally defined metadata framework can enable organizations to enhance trust, interoperability, and security across complex and dynamic data ecosystems. To this end, we rely on standard ontology model to establish a shared security vocabulary, complemented by description logic-based reasoning technologies to address data governance needs. We propose a lightweight and extensible metadata ontology called Metadata Management Ontology (MMO). Our approach seeks to balance conceptual definitions with practical applicability, enabling semantic reasoning, traceability, and cross-system governance. While domain-agnostic in structure, the ontology has been designed with a particular focus on security-sensitive contexts, such as incident investigation, compliance monitoring, and trusted data sharing.

The paper is structured as follows: Section 2 reviews related metadata frameworks and security ontologies. Section 3 introduces the Metadata Management Ontology (MMO), describing its design principles and implementation in Protégé. Section 4 presents the ontology evaluation, including structural metrics and a reasoning scenario involving its integration within an IoT context using SSN ontology. Finally, Section 5 concludes the paper and outlines future research directions.

2 Related Work

Various researchers have explored how to assess the quality of metadata records through multiple dimensions such as accuracy, provenance, consistency, logical coherence, timeliness, accessibility, specificity of subject descriptors, and comprehensiveness [12]. Despite these efforts, the community has yet to agree on a universal framework for defining and measuring these quality aspects. This lack of consensus is largely due to domain-specific requirements and context-driven variations, which necessitate tailored quality criteria depending on the use case [3].

Metadata quality evaluation frameworks have been developed in several works [4]. Nevertheless, these approaches often remain either too domain-specific or too rigid for integration with heterogeneous, dynamic systems. They tend to lack support for metadata that evolve over time or metadata that exist at multiple abstraction levels (e.g., at the level of a data point, a processing step, or a workflow). To tackle this, ontologies, which are sets of concepts used to describe relationships between entities in a machine-understandable format known as RDF, have been developed to enrich metadata expressiveness and enable reasoning over security-relevant properties. For instance, the Unified Cybersecurity Ontology (UCO) [27] aims to unify multiple cybersecurity standards and vocabularies into a coherent OWL-based framework.

Several ontologies tried to capture the essence of data quality but in a superficial light way. For instance, PROV-O [25] provides a standardized vocabulary for modeling and exchanging provenance information across systems. The specification of this ontology serves as the ground work for the implementation of provenance applications across diverse domains. While provenance is a critical dimension of metadata quality, PROV-O does not address other relevant metrics. On the other hand, LIoPY [18] is a legal-compliant ontology that preserves privacy constraints in IoT environments, demonstrating how privacy and legal requirements can be enforced through reasoning mechanisms. Similarly, IdSM-O [13] is an ontology for data sharing management within IoT that aims to preserve security throught the allignent of three existing ontologies: SSN [20], PROV-O, and access control OrBAC ontology. Both papers attempt to infer data quality in the context of data sharing within Internet of Things environments, however they limit their scope to common criteria such as accuracy, frequency, and precision. While the use of metadata as a security and trustworthiness metric is gaining traction, existing approaches often fall short of capturing the full spectrum of quality dimensions. Current models typically rely on domain-specific assumptions and introduce inconsistent evaluation criteria, with limited engagement in comprehensive cross-domain analysis. To the best of our knowledge, there is no widely adopted, domain-agnostic standard for the structured description and assessment of metadata quality. Moreover, although each of these ontologies addresses one or more aspects of metadata, for example, PROV-O focuses on provenance, there is still no unified formal knowledge representation that integrates all of these aspects. Addressing this gap is essential for building interoperable systems where metadata can be reliably used to support security and trust across diverse application contexts.

3 The Metadata Management Ontology (MMO)

Developing an ontology that supports metadata-driven trust, interoperability, and contextual understanding requires a systematic design methodology. Several approaches exist to guide ontology engineering, from manual, process-oriented methods such as TOVE, Ontology Development 101, and the Unified Methodology [26], to more recent semi-automatic approaches using text mining and natural language processing, such as OntoLearn Text-To-Onto [19]. For the Metadata Management Ontology (MMO), we followed MethOntology [7], a well-established, iterative framework that couples rigorous engineering steps with systematic knowledge acquisition. MethOntology organizes development into a structured lifecycle, namely: specification, knowledge acquisition, conceptualisation, formalisation, integration, implementation, and finally evaluation and documentation.

The next subsections first present the conceptual design of the MMO ontology, including class hierarchies, metadata categories, and alignment with existing standards. We then describe its implementation using the Protégé environment, covering editing, validation, and reasoning.

3.1 Ontology design

Specification and knowledge acquisition: The first step in the ontology design was to define the scope and focus. We adopted an

ontological modeling approach to formally structure key concepts and their relationships. In designing the ontology, we considered various metadata types, descriptive, structural, administrative, as well as the tension between static and dynamic metadata. Given the diversity of use cases, we chose to prioritize a descriptive and structural orientation, focusing on the semantics of data objects and their context rather than runtime behavior. Rather than encoding the full complexity of ISO-defined data quality dimensions [10], we leave this aspect to domain-specific extensions. This decision supports interoperability while allowing flexibility for domain-specific extensions.

To better understand the domain, we reviewed metadata-related documentation and standards. The ontology is built around the `Metadata` class, which provides information about the more general `Data` class. The `Data` class is used to represent any information object, such as credentials, attestations, policies, or log events, within security or trust management systems. For example, a field such as `Title` inherits from `Descriptive`, which itself extends the base `Metadata` class, allowing for semantic precision while maintaining generalizability.

Conceptualisation: During this phase, we elicited and organized the domain's key concepts, relations, and constraints. To ensure conceptual coherence and maximize interoperability, the ontology draws inspiration from several established metadata standards, notably Dublin Core (DC) [24] and DCAT [5]. These models provide foundational descriptors for datasets and digital resources, such as title, creator, format, and license, which we have selectively adapted to our use case.

Rather than replicating these standards in their entirety, we opted for a simplified and unified schema that maintains semantic compatibility while reducing complexity. For example, the class `Descriptive` in our ontology encapsulates elements like `Title` and `Description`, echoing the core properties from Dublin Core, but without the full breadth of optional refinements or domain/range constraints. Similarly, `Administrative` aspects such as `AccessRights` or `Provenance` are inspired by DCAT and PROV-O, but represented in a lighter, more modular way that fits dynamic and heterogeneous security scenarios. This alignment strategy allows for semantic mapping and potential future integration with external vocabularies, while ensuring that the ontology remains compact, operational, and easy to extend across domains with minimal overhead.

Initially, we aimed to model data quality directly within the ontology. However, due to the complexity and domain-specific nature of data quality dimensions, we opted to leave quality assessment to the discretion of the ontology's consumers. The model allows for such extensions but does not impose predefined criteria, leaving room for contextual prioritization (e.g., completeness, accuracy, provenance). Rather than modeling metadata as isolated descriptors, we constructed a structured hierarchy based on inheritance. The `Metadata` class is specialized into thematic subclasses, such as `Descriptive`, `Structural`, and `Administrative`, inspired by taxonomies commonly found in metadata literature. These categories can be further refined into more specific concepts, such as `Title` as a subclass of `Descriptive`.

Formalisation: We translated the conceptual model into formal OWL representations. In addition to object properties used to define semantic relationships between individuals (e.g., linking a metadata instance to a log entry or a trust artifact), the ontology also incorporates two key data properties essential for associating literal values, such as strings, numbers, or dates, with metadata elements, enabling more expressive and operational descriptions. Specifically, the ontology defines the following data properties: `label`, a generic textual descriptor attached to any instance of `Metadata`. Through inheritance, this property is applicable to all subclasses and provides a human-readable name or identifier; and `lvalue`, used to associate literal values of any datatype, such as strings, numbers, or dates, with a metadata instance.

These data properties enhance the ontology's expressiveness and usability. They enable fine-grained value assignment and support a range of reasoning and validation tasks, for instance, querying all metadata elements with a numeric threshold, or filtering based on textual labels. By centralizing these properties at the `Metadata` level, and reusing them across subclasses, the model ensures consistency and reduces redundancy in implementation. Each class and property is formally defined to support automated reasoning and rule-based enrichment. The resulting hierarchical organization provides a modular and flexible foundation, well-suited for application across various domains, particularly in systems concerned with digital trust, identity, access control, and policy traceability.

Integration: To avoid duplication and promote interoperability, the ontology reuses and aligns with external standards and vocabularies such as Dublin Core, DCAT and PROV-O, while simplifying and adapting these resources to fit the specific needs of trust management and secure data exchange. The ontology is designed to be extensible: users can introduce new classes, object properties, or data properties tailored to their specific contexts, without disrupting the core structure. This modularity supports domain-specific adaptations while preserving semantic coherence and compatibility with external vocabularies. Moreover, the ontology is designed for seamless integration into existing ontological systems. Any preexisting class from another ontology can support metadata simply by declaring it as a subclass of `Data`. Since `Metadata` is associated with instances of `Data`, this inheritance mechanism enables easy and consistent metadata annotation for external concepts without requiring structural refactoring.

3.2 Ontology Implementation in Protégé

The ontology was developed using the Protégé editor, which facilitated class hierarchy design, property definition, and rule integration in a user-friendly and standards-compliant environment. Figure 1 illustrates the ontology structure within the Protégé interface, showing the inheritance hierarchy among `Metadata` classes (`Descriptive`, `Structural`, `Administrative`,...) and one of their respective subclasses such as `Title`, `Format`, or `AccessRights`.

The ontology was implemented in RDF/XML format, which provides a flexible, graph-based representation for data and metadata. RDF enables linking concepts through triples, making it suitable for integration across distributed and heterogeneous systems [2]. For more expressive semantics, the ontology leverages OWL (Web Ontology Language), which supports class axioms, subclassing, cardinality constraints, and typed data properties, essential for consistent reasoning and interoperability.

To associate literal values with metadata elements, two primary data properties were defined: `label` and `value`. The `value` range was intentionally left untyped (i.e., without a fixed `rdfs:range`) to maximize flexibility. However, in cases where type enforcement is necessary (e.g., ensuring a value is an integer or a date), Protégé allows the manual specification of datatype restrictions using OWL constructs such as `xsd:integer`, `xsd:string`, or `xsd:date`. These constraints can be applied via class-level axioms or property range restrictions, depending on the precision required for a specific use case. The full ontology is available as open-source and maintained in a public GitLab repository[1].

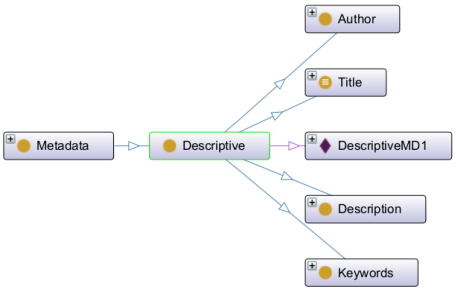

Figure 1: Descriptive Metadata classes in Protégé

4 Ontology Evaluation

Ontology evaluation aims to compare an ontology against specification requirements, such as competency questions that are meant to be solvable by the developed ontology. We evaluate three main aspects: the correctness of the infered results, the integration of MMO with existing ontologies, and the scalability performance of the model.

4.1 Correctness Validation

According to [28], there are various criteria for ontology evaluation: accuracy, which ensures alignment between axioms and expert knowledge; adaptability, reflecting the ontology's extensibility and ability to support diverse domain tasks; completeness, which assesses how well the ontology captures the full scope and nuances of the domain; computational efficiency, referring to the performance of reasoning and querying operations; and conciseness, which emphasizes the exclusion of redundant or unnecessary concepts to maintain a streamlined and non-redundant structure. Direct measurement of the mentioned criteria is difficult. Therefore, we would be looking for metrics to express those requirements and assess both the ontology correctness and ontology quality. In this context, we create a matrix between the previously mentioned criteria and OntoMetrics proposed in [15].

We will evaluate MMO based on four major metrics: schema, knowledge base, class, and graph. Respectively, those metrics aim to address the design and richness of an ontology, instances distribution within it, examine the classes and relations, and finally measures to calculate the structure of the ontology. Firstly, we evaluate the ontology's correctness, which includes accuracy, completeness, and conciseness. Secondly, we assess the ontology's quality, which encompasses computational efficiency and adaptability.

Table 1 shows the evaluation document that details MMO calculated quality metrics for ontology validation. A positive correlation is between the OntoMetrics criteria and Vrandečić ontology evaluation metrics. The evaluation metrics confirm that MMO is structurally sound and ready for deployment. Attribute richness (0.34) balances semantic expressiveness with simplicity, while inheritance richness (1.08) and relationship richness (0.44) indicate a well-distributed hierarchy and meaningful connectivity. Completeness is supported by an axiom/class ratio of 7.45, and conciseness is shown through a class children count of 4 and inheritance depth of 5. Efficiency metrics, angledness of 0.6, 32 paths and a maximum depth of 74, demonstrate scalability without heavy reasoning costs. With an average population of 0.3 and class richness of 0.25, the ontology remains lightweight. Finally, adaptability is validated by a sibling cardinality of 20 and 16 leaf nodes, confirming MMO's ability to incorporate new concepts without disrupting its structure.

4.2 Case Study: Internet of Things (IoT)

We illustrated our ideas in the Internet of Things (IoT) context, but our ontology is agnostic and can be applied in other domains. Thus, we describe the following simple scenario. In a smart city context, sensors distributed across various neighborhoods collect environmental data such as CO2 concentration, with each `ssn:Observation` enriched using the MMO ontology to enhance semantic context. Metadata annotations include `mmo:Frequency`, indicating how often the observation is updated.

To enable semantic classification based on update frequency, we define a SWRL rule that infers properties of the stimulus detected by the sensor. The rule identifies an instance of `ssn:Observation` linked to a metadata element of type `mmo:Frequency` and checks whether the associated `mmo:value` exceeds a threshold of `0.7`. When this condition is satisfied, the rule traces the observation back to the sensor that generated it (`ssn:madeBySensor`), and further to the stimulus that the sensor detects (`ssn:detects`). Finally, it infers that the stimulus is a proxy for a high-frequency property using the SSN relation `ssn:isProxyFor`.

The rule is expressed as follows:

```
ssn:Observation(?obs) ∧ mmo:hasMetadata(?obs,
?meta) ∧ mmo:Frequency(?meta) ∧
mmo:value(?meta, ?v) ∧ swrlb:greaterThan(?v,
0.7) ∧ ssn:madeBySensor(?obs, ?sensor) ∧
ssn:detects(?sensor, ?stimulus)
→ ssn:isProxyFor(?stimulus, :HighFreqProperty)
```

To validate the semantic interoperability between the SSN and MMO ontologies, we used the Protégé editor with the Pellet reasoner. After importing both ontologies, we instantiated a sample observation and annotated it with relevant metadata. The individual `:CO2_obs` was declared as an instance of `ssn:Observation`, and linked to a metadata instance `:CO2_meta_frequency` of type `mmo:Frequency` with a numerical value of `0.9` (typed as `xsd:float`). This setup was encoded as follows:

[1] https://git.univ-pau.fr/munier/mmo

	Ontology Correctness			Ontology Quality	
	Accuracy	**Completeness**	**Conciseness**	**Computational Efficiency**	**Adaptability**
Schema metrics					
Attribute richness	0.34	–	–	–	–
Inheritance richness	1.08	–	–	–	–
Relationship richness	0.44	–	–	–	–
Axiom/class ratio	–	7.45	–	–	–
Graph metrics					
Absolute sibling cardinality	–	–	–	–	20
Absolute depth	–	–	–	74	–
Total number of paths	–	–	–	32	–
Class metrics					
Class inheritance richness	5	–	–	–	–
Class children count	–	–	4	–	–
Knowledgebase metrics					
Average population	–	–	–	0.3	–
Class richness	0.25	–	–	–	–

Table 1: MMO evaluation document

```
:CO2_obs rdf:type ssn:Observation .
:CO2_meta_frequency rdf:type mmo:Frequency ;
        mmo:value "0.9" xsd:float .
:CO2_obs mmo:hasMetadata :CO2_meta_frequency .
:CO2_obs ssn:madeBySensor :CO2_sensor .
:CO2_sensor ssn:detects :CO2_stimulus .
```

A SWRL rule was defined to classify observations by update frequency and infer system-level effects. The reasoning engine checks if an `ssn:Observation` is linked to a `mmo:Frequency` with `mmo:value` above `0.7`. Since `:CO2_obs` has a frequency value of `0.9`, and its sensor `CO2_sensor` detects `CO2_stimulus`, the reasoner infers that `CO2_stimulus` is serving as a proxy for `:HighFreqProperty`. Upon reasoning, the engine correctly classified `:CO2_stimulus` under this rule and inferred the right output, confirming the rule's validity and successful cross-ontology reasoning.

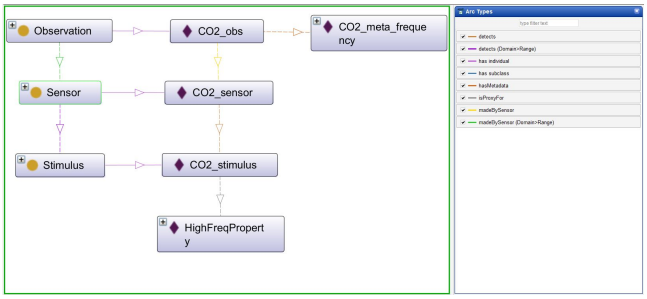

Figure 2: Result after launching the inference engine

Althought not complex, this use case allows us to successfully validate the reasoning process in order to check logical consistency and perform inference, confirming that our ontology is both semantically well-formed and logically coherent. This process also demonstrates the correct integration of our model with the SSN ontology. Thanks to its lightweight, extensible, and interoperable design, MMO can be easily incorporated into intelligent systems or other OWL ontologies. By simply subclassing `Data`, any external class can inherit the metadata properties defined in our ontology, enabling efficient semantic enrichment in complex, heterogeneous environments without introducing heavy dependencies or domain-specific constraints. In addition, MMO can be used in more security-sensitive contexts, such as regulatory compliance, where metadata traceability ensures accountability, or incident response, and where fine-grained provenance analysis facilitates root-cause investigation.

4.3 Scalability

We conduct some experiment explore the behaviour of MMO in large-scale settings with a higher number of observations and metadata instances. To do so we use OWLAPI and Java to measure and compare the performance response using SWRL API. To this end, we evaluate whether the computational time of the reasoning is acceptable by performing several tests on the same rule mentioned before, while increasing the number of observations from 1000 to 120000. Each observation have three types of metadata: description, frequency, and modification date. Thus, we perform an experiment to measure the time needed to check and select all the stimulus who are proxy for a high frequency property. Fifty per cent of each number of observation returns false, while the other half returns true to the query.

Figure 3 depicts the performance response while increasing obserations number from 1000 to 120000. The processing time varies from 0.6 to 11 seconds. The quasi linearity property behind these results means that a better computer system setting would obtain a lower processing time.

5 Conclusion and Future Work

Recognizing the importance of capturing data features and manipulation descriptions in distributed environments, in this paper we

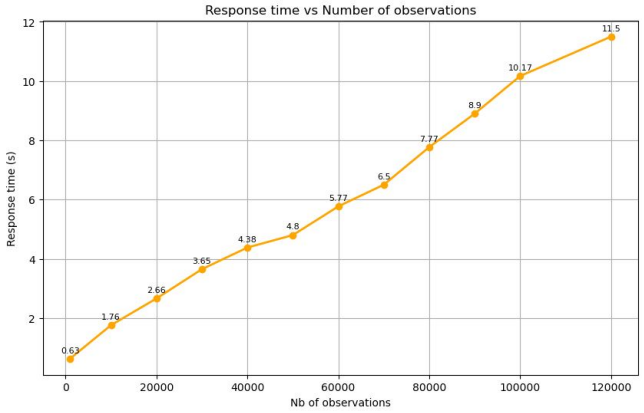

Figure 3: Response time vs number of observations

advocate for the role of metadata towards this goal. We introduced a lightweight ontology designed to unify the representation of metadata across heterogeneous systems. We validated the ontology's semantic coherence and logical consistency through structural metrics and rule-based reasoning. Its successful integration into established frameworks, such as the W3C SSN ontology, demonstrates its practical interoperability.

As future work, a key use case of our approach is AI, where assessing dataset quality is critical to ensuring the reliability of model outputs. Additionally, metadata could play an essential role in AI explainability by documenting how data is processed prior to model training, thereby ensuring trust in the resulting models. Authors in [16] has explored this vision, focusing solely on data provenance as a metadata feature. In contrast, our approach hope to broadens this perspective by encompassing a wider range of metadata types, using a formal standard. In future work, we plan to extend the ontology with explicit support for distinguishing between *static* and *dynamic* metadata. This will allow for better modeling of evolving contexts, particularly in time-sensitive or real-time environments such as cybersecurity systems. We also intend to reinforce the integration of provenance and trust mechanisms with PROV-O to ensure traceability, source accountability, and policy enforcement. In security-critical domains, our ontology provides a foundational layer for trust-aware metadata that supports transparency, governance, and resilience.

References

[1] Saeed Hamood Alsamhi, Raushan Myrzashova, Ammar Hawbani, and Santosh Kumar. 2024. Federated learning meets blockchain in decentralized data-sharing: Healthcare use case. *IEEE Internet of Things Journal* (2024).

[2] Farshad Badie. 2020. A formal ontology for conception representation in terminological systems. *Reasoning: Logic, cognition, and games/ed. by M. Urbanski, T. Skura and P. Lupkowski. London: College Publications* (2020).

[3] Cristóbal Barba-González, Ismael Caballero, Ángel Jesús Varela-Vaca, José A Cruz-Lemus, María Teresa Gómez-López, and Ismael Navas-Delgado. 2024. BIGOWL4DQ: Ontology-driven approach for Big Data quality meta-modelling, selection and reasoning. *Information and Software Technology* 167 (2024), 107378.

[4] Gustavo Candela, Pilar Escobar, Rafael C Carrasco, and Manuel Marco-Such. 2022. Evaluating the quality of linked open data in digital libraries. *Journal of Information Science* 48, 1 (2022), 21–43.

[5] Kaifang Dong, Yifan Liu, Fuyong Xu, and Peiyu Liu. 2023. DCAT: combining multisemantic dual-channel attention fusion for text classification. *IEEE Intelligent Systems* 38, 4 (2023), 10–19.

[6] Widad Elouataoui, Saida El Mendili, and Youssef Gahi. 2024. Active Metadata and Machine Learning based Framework for Enhancing Big Data Quality. In *Proceedings of the 7th International Conference on Networking, Intelligent Systems and Security*. 1–8.

[7] Mariano Fernández-López, Asunción Gómez-Pérez, and Natalia Juristo Juzgado. 1997. Methontology: from ontological art towards ontological engineering. (1997).

[8] Pau Ferrer-Cid, Jose M Barcelo-Ordinas, and Jorge Garcia-Vidal. 2025. A review of graph-powered data quality applications for IoT monitoring sensor networks. *Journal of Network and Computer Applications* (2025), 104116.

[9] Istvan Haller, Erik Van Der Kouwe, Cristiano Giuffrida, and Herbert Bos. 2016. METAlloc: Efficient and comprehensive metadata management for software security hardening. In *Proceedings of the 9th European Workshop on System Security*. 1–6.

[10] ISO. 2022. Data quality. https://www.iso.org/standard/81745.html ISO 8000 series.

[11] Ruchun Jia, Jianwei Zhang, and Yi Lin. 2024. Machine Learning Security Defense Algorithms Based on Metadata Correlation Features. *Computers, Materials & Continua* 78, 2 (2024).

[12] Péter Király. 2019. *Measuring metadata quality*. Ph. D. Dissertation. Georg-August-Universität Göttingen.

[13] Nouha Laamech, Manuel Munier, and Congduc Pham. 2022. IdSM-O: An IoT data sharing management ontology for data governance. In *Proceedings of the 14th International Conference on Management of Digital EcoSystems*. 88–95.

[14] Stefano Carlo Lambertenghi and Andrea Stocco. 2024. Assessing quality metrics for neural reality gap input mitigation in autonomous driving testing. In *2024 IEEE Conference on Software Testing, Verification and Validation (ICST)*. IEEE, 173–184.

[15] Birger Lantow. 2016. Ontometrics: Putting metrics into use for ontology evaluation. In *KEOD*. 186–191.

[16] Luca Lauro, Pasquale Leonardo Lazzaro, Marialaura Lazzaro, Paolo Missier, and Riccardo Torlone. 2024. An LLM-guided Platform for Multi-Granular Collection and Management of Data Provenance. (2024).

[17] Jieh-Sheng Lee. 2020. Controlling Patent Text Generation by Structural Metadata. In *Proceedings of the 29th ACM International Conference on Information & Knowledge Management (CIKM '20)*. ACM, 3241–3244. https://doi.org/10.1145/3340531.3418503

[18] Faiza Loukil, Chirine Ghedira-Guegan, Khouloud Boukadi, and Aicha Nabila Benharkat. 2018. LIoPY: A legal compliant ontology to preserve privacy for the Internet of Things. In *2018 IEEE 42nd annual computer software and applications conference (COMPSAC)*, Vol. 2. IEEE, 701–706.

[19] Alexander Maedche and Steffen Staab. 2000. The text-to-onto ontology learning environment. In *Software Demonstration at ICCS-2000-Eight International Conference on Conceptual Structures*, Vol. 38. Citeseer, 890–930.

[20] Musa Milli, Mehmet Milli, Sanaz Lakestani, Özlem Aktaş, et al. 2023. Semantic-based anomaly detection in laboratory environments using SOSA/SSN sensor ontology frameworks. *Pamukkale University Journal of Engineering Sciences* 29, 4 (2023), 357–369.

[21] Thanh Linh Nguyen, Lam Nguyen, Thong Hoang, Dilum Bandara, Qin Wang, Qinghua Lu, Xiwei Xu, Liming Zhu, and Shiping Chen. 2025. Blockchain-empowered trustworthy data sharing: Fundamentals, applications, and challenges. *Comput. Surveys* 57, 8 (2025), 1–36.

[22] National Information Standards Organization (NISO). 2007. A framework of guidance for building good digital collections. In *https://www.niso.org/sites/default/files/2017-08/framework3.pdf*. 1–8.

[23] Javier Nogueras-Iso, Javier Lacasta, Manuel Antonio Ureña-Cámara, and Francisco Javier Ariza-López. 2021. Quality of metadata in open data portals. *IEEE Access* 9 (2021), 60364–60382.

[24] Jung-ran Park and Eric Childress. 2009. Dublin Core metadata semantics: An analysis of the perspectives of information professionals. *Journal of Information Science* 35, 6 (2009), 727–739.

[25] Tim Prudhomme, Giacomo De Colle, Austin Liebers, Alec Sculley, Peihong "Karl" Xie, Sydney Cohen, and John Beverley. 2025. A semantic approach to mapping the Provenance Ontology to Basic Formal Ontology. *Scientific Data* 12, 1 (2025), 282.

[26] Abdul Sattar, Ely Salwana Mat Surin, and Ahmadl. 2020. Comparative analysis of methodologies for domain ontology development: A systematic review. *International Journal of Advanced Computer Science and Applications* (2020).

[27] Zareen Syed, Ankur Padia, M Lisa Mathews, Tim Finin, Anupam Joshi, et al. 2016. UCO: A unified cybersecurity ontology. In *Proceedings of the AAAI Workshop on Artificial Intelligence for Cyber Security*. 195–202.

[28] Denny Vrandečić. 2009. Ontology evaluation. In *Handbook on ontologies*. Springer, 293–313.

[29] Marco Zappatore, Antonella Longo, Angelo Martella, Beniamino Di Martino, Antonio Esposito, and Serena Angela Gracco. 2023. Semantic models for IoT sensing to infer environment–wellness relationships. *Future Generation Computer Systems* 140 (2023), 1–17. doi:10.1016/j.future.2022.10.005