Semantic Interoperability in Smart Cities: A Distributed Architecture Based on Ontologies, Plug-and-Play AI, and BlockDAG

Tatiana Caballero, Derlis O. Gregor*
Faculty of Engineering, National University of Asuncion, San Lorenzo, Paraguay
Emails: t.caballero@fiuna.edu.py, dgregor@ing.una.py

***Abstract*— Smart City platforms increasingly depend on sensing devices and distributed services. The implementation of data-driven mobility systems and infrastructure management has not solved the problem of semantic interoperability. The problem comes from data models in middleware solutions and from the lack of support for devices connecting on their own. Existing approaches often struggle to scale under high data rates and dynamic environments. We presents a work-in-progress distributed architecture that combines semantic technologies, Plug-and-Play Artificial Intelligence (PnP-AI), and a Directed Acyclic Graph (DAG)-based synchronization layer that aims to enable decentralized semantic interoperability in Smart City scenarios. The system applies an Intelligent Transportation System ontology to perform semantic labeling of sensor observation, Shapes Constraint Language (SHACL) shapes for local semantic validation, and a lightweight PnP-AI agent with a Decision Tree classifier to perform automatic device categorization during onboarding. The system design allows nodes to perform processing operations and validation tasks while semantically validated events are propagated through a DAG-based ledger without relying on centralized middleware. Preliminary experimental results from a simulated multi-node environment showed that end-to-end latency remained under 15 ms and throughput reached more than 100 events per second while all messages successfully passed SHACL validation. These results show that the architecture can improve scalability and resilience compared to centralized middleware systems. Current efforts are directed toward expanding the test scope, incorporating security and privacy mechanisms, and demonstrating the approach in more complex urban mobility scenarios.**

Index Terms— Smart Cities, Semantic Interoperability, Ontologies, SHACL, BlockDAG, Plug-and-Play AI, Distributed Middleware, Intelligent Transportation Systems.

I. Introduction

Smart Cities operate through extensive networks of diverse sensing equipment and spread-out urban services, which produce continuous real-time data streams for traffic control, mobility statistics, and public security needs. The Smart City deployments experience fragmentation because different systems use incompatible data formats and communication protocols and semantic models, which prevents them from working together at scale [1].

The solution to this problem lies in semantic interoperability, which creates a common understanding across different systems by using formal descriptions such as RDF, OWL, and SHACL. The technologies allow for automated validation and sensor data improvement through environmental context processing, which operates in urban environments [2]. The current solutions for IoT systems rely on centralized semantic middleware, which creates performance limitations, system failures, and a limited ability to handle changing IoT environments.

Research has used blockchain technology to study decentralization because this system creates trust while enabling better data sharing and enhanced audit capabilities for Smart Cities [3]. The linear design of blockchains imposes performance constraints that prevent real-time operation, as it limits both system speed and response time. The DAG-based ledger system solves these problems by allowing multiple validation operations to run simultaneously while supporting large-scale IoT systems [4], but it does not manage semantic interoperability within its system.

The Plug-and-Play AI (PnP-AI) system has been developed to reduce manual setup requirements by enabling self-service device and service identification [5]. Research studies demonstrate that AI and distributed ledgers can merge into a new system, which shows promise for building dependable Smart City infrastructure [6]. The current methods use AI, semantics, and distributed ledgers as separate elements that do not fully integrate with each other.

To the best of our knowledge, no existing architecture jointly provides decentralized semantic validation, autonomous device onboarding, and DAG-based event synchronization. This paper addresses that gap with the following contribution: we propose a distributed semantic interoperability architecture that integrates an ITS ontology with SHACL-based local validation, a lightweight PnP-AI module for autonomous device onboarding, and a DAG-based synchronization layer. Each peer node performs the full semantic processing pipeline independently and propagates only validated event summaries through the DAG ledger, avoiding centralized middleware while maintaining low latency and stable throughput in simulation.

II. Related Work

A. Semantic Interoperability in Smart Cities

Smart City platforms require semantic interoperability as their basic operational need because different devices and services need to share data with a common understanding. The research uses RDF and OWL along with ontologies to

*Corresponding author: dgregor@ing.una.py

create models that describe urban entities and sensor-based observations according to [2]. Most deployment systems depend on centralized semantic middleware, which creates performance restrictions and longer response times, and vulnerable system entry points when dealing with high data transfer rates [1].

B. Ontology-Based Approaches in Intelligent Transportation Systems

The application domain of Intelligent Transportation Systems (ITS) has reached full maturity for semantic modeling. Researchers have developed ontologies that enable the creation of reusable, interoperable representations of traffic sensors, vehicles, and their associated mobility events. The methodology presented in [7] offers systematic procedures for ITS ontology development, which has become the standard approach in the field.

The most ontology-based ITS frameworks use centralized data processing pipelines, which restrict their use in decentralized systems that need local validation and autonomous operation.

C. Blockchain and DAG-Based Architectures for Smart Cities

Research has been conducted to use blockchain technology for Smart City infrastructure decentralization and enhanced trust and auditing capabilities [3]. The sequential consensus process in linear blockchains creates performance issues which prevent them from working for real-time urban sensing applications because they experience both high latency and restricted throughput.

The DAG-based distributed ledgers system solves these problems through its parallel validation capability, which leads to enhanced scalability according to [4]. The security and robustness issues that DAG architectures experience during low-traffic periods receive most research attention, but scientists have not studied semantic interoperability enough.

D. AI-Driven Autonomous Device Onboarding

The Plug-and-Play AI system conducts self-operating device identification and setup processes, which allow it to automatically connect network devices across different network environments [5]. The deployment of lightweight machine learning models at edge locations serves two purposes, which include device capability inference and processing pipeline selection. The combination of AI with distributed ledger technology systems creates improved automated systems that maintain their operational stability in Smart Cities, according to [6].

The current solutions lack integration between AI-based onboarding systems and semantic validation and decentralized coordination systems.

E. Research Gap

The research investigates semantic interoperability and ontology-based modeling, distributed ledgers, and AI-driven automation as three distinct individual subjects. Centralized middleware limits scalability, blockchain-based approaches introduce latency, and DAG-based ledgers rarely incorporate semantic processing. The architecture proposed in this paper is motivated by this gap, as detailed in Section I.

III. Methodology and Architecture

The methodology of this study implements a work-in-progress approach to assess decentralized semantic interoperability functionality in Smart City environments. The solution proposes to use semantic technologies together with autonomous device classification and DAG-based synchronization, which will operate through a peer-to-peer execution system. The system operates through independent nodes, which perform their entire processing pipeline without needing any centralized middleware components. Figure 1 illustrates the end-to-end workflow performed at each peer node.

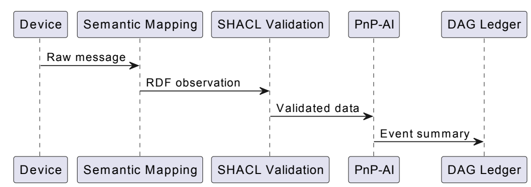

Fig. 1. Decentralized semantic interoperability workflow executed locally at each peer node, including autonomous device classification, ontology-based RDF mapping, SHACL validation, and DAG-based event dissemination.

A. Overall Architecture

The full processing pipeline runs on each peer node, which operates independently without using centralized middleware components. The system consists of four fundamental modules, which serve as its fundamental operational framework:

- **The Semantic Mapping Module** transforms device messages into RDF representations through an ITS-based ontology.
- **The SHACL Validation Module** performs local constraint-based validation of RDF messages to check their structural and semantic accuracy.
- The system includes a **Plug-and-Play AI Module** which performs device and observation classification to enable self-configuration and semantic setup of devices.
- **The DAG Synchronization Module** sends validated event summaries to a distributed DAG-based ledger system.

The system runs nodes as separate local processes, which create a peer-to-peer deployment environment. The system design allows researchers to study system scalability, latency, and validation performance without creating any performance limitations throughout the entire system architecture.

B. Adversary Model and Threat Assumptions

The prototype operates with a minimal security threat model, which works for first deployments of decentralized systems. The semantic layer performs node-local SHACL validation to stop incorrect RDF data entries and absent properties, and attempts to evade constraints while it blocks all invalid messages before they reach other parts of the system. The ledger layer identifies duplicate events and replayed events through its use of cryptographic hash references, while stronger guarantees against Sybil attacks, spam flooding, and low-traffic DAG vulnerabilities are considered future work.

C. Semantic Representation and Validation

The system uses an existing ITS ontology to establish a framework that includes sensor data and observation information, and mobility-related elements. Raw device messages are converted into RDF graphs using RDFLib and the SOSA (Sensor, Observation, Sample, and Actuator) observation model. The system maintains semantic correctness through PySHACL, which applies SHACL shapes to detect both invalid and inconsistent data at the node level.

The SHACL shapes used in this implementation establish rules that control sensor identification and observation types and data types, and time-based data attribution. The prototype does not perform TBox reasoning or ontology-based rewriting in order to keep validation overhead constant at each individual node. The implementation of advanced SHACL+OWL integration methods would improve semantic completeness, but it would generate new processing difficulties that are left for future work.

D. Plug-and-Play for Autonomous Onboarding

To support autonomous device integration in a decentralized setting, a lightweight Plug-and-Play Artificial Intelligence (PnP-AI) module based on a Decision Tree classifier is introduced. In this context, PnP-AI refers to the capacity of the system to automatically identify, classify, and configure new devices upon their first connection, without manual intervention. The module performs basic semantic classification of new devices and services at its initial stage instead of conducting detailed device identification processes. The present work-in-progress implementation includes two broad categories, which are Intelligent Vehicles and Intelligent Infrastructures that serve practical purposes in Intelligent Transportation Systems operations.

The classifier is trained in isolation using a synthetic dataset derived from the existing ITS ontology, where each training instance corresponds to an ontology class, and labels represent the associated high-level semantic category. The system extracts all input features directly from ontology-level data, which includes class identifiers and hierarchical depth, and parent-child connection information. The system design uses semantic metadata for onboarding operations, which operate independently from sensor data and specific domain information.

The dataset undergoes stratified splitting to create training and test subsets, which prevents data leakage from occurring. Class imbalance present in the ontology-derived dataset is addressed only in the training subset through oversampling of the minority class. The model complexity receives control through two mechanisms, which restrict tree depth and minimum leaf size to prevent overfitting while maintaining interpretability. The trained model gets converted into a serialized format, which peer nodes can deploy locally for decentralized device classification without requiring central coordination.

The classification process does not modify the organization of RDF data or the sensor-based information storage. The system enables users to choose ontology modules and SHACL shapes, which they can use for conducting local semantic validation operations. The system enables nodes to establish their own semantic rules based on context, which they can enforce without losing their decentralized nature.

The evaluation of generalization behavior occurs through learning curve analysis, which shows in Figure 2 that training and validation accuracy match as the number of samples grows, which proves weak overfitting and stable generalization. The system maintains its correct operation and ledger integrity because misclassification does not affect the system's functionality. The system will choose a non-optimal SHACL shape during shape selection when it makes an incorrect class assignment, which results in elevated local rejection rates, but the system will prevent all semantically invalid messages from entering the DAG ledger. Reclassification can be triggered by repeated SHACL failures or detected inconsistencies, or handled through a fallback generic constraint set. The system produces incorrect results, which impacts onboarding time duration and acceptance success rates, but does not affect the overall system safety or global performance.

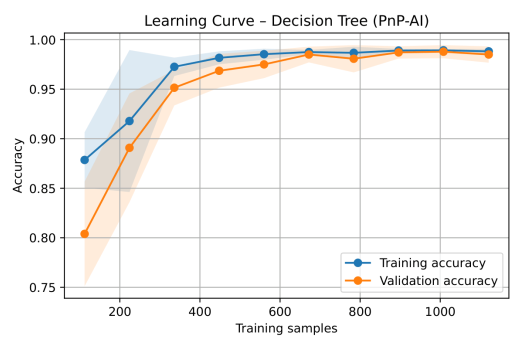

Fig. 2. Learning curve analysis of the PnP-AI Decision Tree classifier, showing convergence between training and validation accuracy as the number of samples increases.

E. DAG-Based Event Synchronization

After semantic validation, event summaries are published to a Directed Acyclic Graph (DAG)-based ledger. The system uses multiple previous events in each operation to perform parallel validation at high speeds while allowing many operations to run simultaneously without requiring any sequence of events. The system achieves lower latency and reduced computational requirements because of its design structure.

The DAG contains only hashed summaries together with basic metadata, which maintains data integrity while allowing audit tracking without showing actual sensor data. The system operates with a basic DAG structure that resembles

Tangle systems at the prototype level. Tip selection and confirmation are managed through basic heuristic methods that substitute for formal consensus procedures. The current implementation provides integrity and auditability of hashed event summaries; features such as Sybil resistance, spam mitigation, and robust finality guarantees are considered future work.

F. Experiment Setup and Metrics

Experimental evaluation is conducted through a controlled simulation of a peer-to-peer Smart City environment. Each peer node is executed as an independent local process, emulating decentralized deployment without shared middleware components, and exchanges semantic messages through a simulated DAG-based synchronization layer.

End-to-end latency is measured from the reception of a semantic message at a node to the insertion of the corresponding validated event summary into the DAG. This metric includes RDF mapping, SHACL validation, PnP-AI inference, and local DAG insertion, while explicitly excluding network propagation delays, cryptographic signing, and ledger-level consensus finality, which are scoped as future work.

Experiments are executed on a single host machine equipped with an Intel Core i7 CPU (8 logical cores), 16 GB of RAM, and a Windows-based operating system. Each peer node runs without explicit CPU pinning using process isolation. Messages consist of small RDF graphs (10–100 triples) and compact event summaries. Performance is evaluated using mean and 95th-percentile end-to-end latency, throughput (events/s), SHACL validation overhead, and semantic conformance rate (Table II).

IV. Preliminary Results

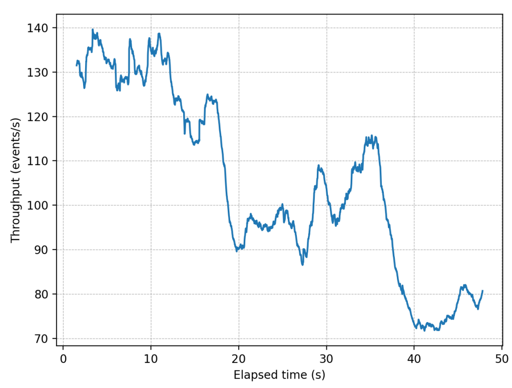

Fig. 3. Throughput evolution over time for the DAG-based architecture under simulated load.

Figure 3 illustrates the temporal evolution of system throughput during the experiment. The DAG-based architecture maintains operation at more than 100 events per second throughout most of the execution period because it manages message scheduling and local validation delays. The system operates stably because parallel event validation functions with missing global consensus mechanisms to manage continuous system traffic. The reported throughput shows the complete duration of semantic processing, which includes RDF mapping, SHACL validation, and Plug-and-Play AI inference and local DAG insertion operations instead of showing the speed of basic centralized data ingestion.

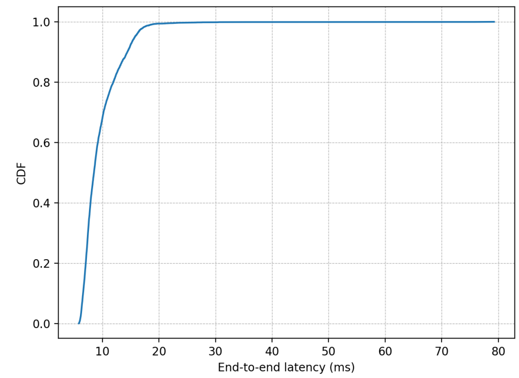

Fig. 4. Cumulative distribution function (CDF) of end-to-end latency for semantic event processing.

The cumulative distribution of end-to-end latency for semantic event processing appears in Figure 4. More than 90% of the events are processed with latency below 15 ms, while the tail latency remains bounded below 80 ms. The system produces expected results according to the distribution because DAG-based synchronization with local semantic validation maintains low delays when multiple nodes execute the system.

Table I compares the proposed architecture against reference values reported in the literature for centralized semantic middleware [1] and blockchain-based Smart City systems [3], [6]. The Middleware and Blockchain columns report performance ranges documented in prior work under production or near-production conditions; the DAG column reports results obtained experimentally in this study under controlled simulation. The proposed architecture achieves mean latency of 9.56 ms and throughput above 100 events/s, comparing favorably against both reference baselines.

TABLE I
PRELIMINARY PERFORMANCE METRICS

Metric	Middleware	Blockchain	DAG (proposed)
Avg. latency (ms)	250	1500	9.56
Throughput (events/s)	300	50	105
Validation overhead	High	Medium	Low

† Reference values of Middleware and Blockchain are from literature

The quantitative results from the simulated multi-node deployment of the proposed architecture appear in Table II. The system processes 5,000 semantic messages between three peer nodes while keeping end-to-end latency below 10 ms and 95th-percentile latency at or below 16 ms. The system operates at a constant throughput rate when processing more than 100 events per second because semantic validation operations do not cause major performance delays.

The processed messages meet all SHACL constraints which results in a semantic conformance rate of 1.0. The results show that DAG-based synchronization for local semantic validation operates at high speeds, which do not create performance-degrading delays in the system.

TABLE II
SUMMARY OF EXPERIMENTAL RESULTS FOR THE DAG-BASED

Metric	Value	Unit
Messages processed	5000	–
Number of nodes	3	–
Mean latency	9.56	ms
95th-percentile latency	15.74	ms
Throughput	104.6	events/s
SHACL overhead (mean)	6.15	ms
Semantic conformance rate	1.00	ratio

Table III summarizes the classification performance obtained on a held-out test set. The model achieves 0.96 overall accuracy while it maintains high precision, recall, and F1-score values for both *Intelligent Vehicles* and *Intelligent Infrastructures*. These two categories correspond directly to the branches of the ITS ontology used for SHACL shape selection during onboarding: a correctly classified device is assigned the appropriate constraint set, ensuring that the RDF observations it subsequently generates are validated against semantically relevant rules before being inserted into the DAG ledger. The learning curve analysis demonstrates that the chosen model configuration produces reliable results without any major overfitting problems.

TABLE III
CLASSIFICATION PERFORMANCE OF THE PLUG-AND-PLAY AI DECISION TREE MODEL

Class	Precision	Recall	F1-score	Support
Intelligent Infrastructures	0.99	0.95	0.97	176
Intelligent Vehicles	0.91	0.99	0.95	90
Accuracy		0.96		

V. DISCUSSION AND FUTURE WORK

The system design enables decentralized semantic interoperability by maintaining high data transfer rates and quick system responses. The system operates at high efficiency by performing local SHACL validation and DAG-based event dissemination, eliminating the need for centralized middleware to address Smart City platform scalability and resilience issues.

The PnP-AI module demonstrates excellent classification results through its implementation of ontology-based features together with a compact Decision Tree model, which proves effective for initial autonomous system enrollment processes. The use of synthetic data alongside basic semantic categories limits the method's effectiveness across different situations, requiring additional testing.

The security and governance system of the current prototype uses a basic threat model that prioritizes system architectural development. SHACL validation helps prevent semantic input errors, but researchers need to develop security protocols that protect against adversarial attacks that affect both semantic data and ledger systems.

The research will continue to evaluate its performance using actual data streams, which will also study Smart City applications beyond their current scope. In addition, the research will examine different methods for system onboarding, develop more complex semantic rules for data interpretation, and study two main topics: developing more effective SHACL engines and methods to transform ontologies for improved performance. Last but not least, the research will continue to develop solutions for identity management and access control, as well as disclosure control systems, which will be tested through a distributed testbed that simulates actual network environments.

VI. ACKNOWLEDGMENT

This publication has been prepared with the support of the National Council of Science and Technology (CONACYT), Paraguay. The content is the sole responsibility of the authors and under no circumstances should it be considered to reflect the views of the CONACYT. The Project PINV01-24, "Interoperabilidad de Equipos y Dispositivos con capacidad Plug and Play aplicado a Sistemas de Transportes Inteligentes: Un paso hacia las Smart Cities," from which this work originates, is co-financed by the CONACYT with support from the Fund for Excellence in Education and Research (FEEI). The authors also express their gratitude for the inter-institutional collaboration between the Faculty of Engineering of the National University of Asuncion (FIUNA), Paraguay, and the Higher Technical School of Engineering (ETSI) of the University of Seville (US), Spain, for the technical and academic support provided throughout the development of the project.

REFERENCES

[1] S. Biswas, Z. Yao, L. Yan, A. Alqhatani, A. K. Bairagi, F. Asiri, and M. Masud, "Interoperability in smart cities: Benefits, challenges and architectural perspectives," *Electronics*, vol. 12, no. 1036, 2023.

[2] A. Pliatsios, K. Kotis, and C. Goumopoulos, "A systematic review on semantic interoperability in the ioe-enabled smart cities," *Internet of Things*, vol. 22, p. 100754, 2023.

[3] F. Zahrae Chentouf, M. El Alami Hasoun, and S. Bouchkaren, "Blockchain for sustainable smart cities: Motivations and challenges," *Computer Science and Mathematics Forum*, vol. 10, no. 2, 2025.

[4] Y. Bai, S. Lee, and S.-H. Seo, "A survey on directed acyclic graph-based blockchain in smart mobility," *Sensors*, vol. 25, no. 1108, 2025.

[5] I. Al Ridhawi, S. Otoum, M. Aloqaily, and A. Boukerche, "Generalizing ai: Challenges and opportunities for plug and play ai solutions," *IEEE Network*, vol. 35, 2020.

[6] A. Sharma, E. Podoplelova, A. Tselykh, and A. Tselykh, "Sustainable smart cities: Convergence of artificial intelligence and blockchain," *Sustainability*, vol. 13, no. 13076, 2021.

[7] D. Gregor, S. Toral, T. Ariza, F. Barrero, R. Gregor, J. Rodas, and M. Arzamendia, "A methodology for structured ontology construction on intelligent transportation systems," *Computer Standards and Interfaces*, vol. 47, 2015.