Knowledge Graph-Driven Policy Enforcement in Urban Dataspace

Dessislava Petrova-Antonova[1]*, Paolo Pareti[2]*, Petar Tomov[1], Semih Yumusak[2], Christopher Maidens[2], Shanshan Jiang[3], George Konstantinidis[2], and Dumitru Roman[3,4]

[1] GATE Institute, Sofia University, Sofia, Bulgaria
[2] University of Southampton, Southampton, UK
[3] SINTEF AS, Norway
[4] Bucharest University of Economic Studies, Romania
Contact: dessislava.petrova@gate-ai.eu

Abstract. With the diffusion of interconnected sensors, systems, and stakeholders, smart cities are evolving into complex, data-intensive environments where urban data has become a critical asset. New challenges emerge in urban dataspaces regarding enforcing multi-actor, heterogeneous, dynamic and real-time policies. To explore the potential of applying Knowledge Graph (KG) technologies in the urban context, we identified a comprehensive list of real-world policy challenges related to data governance through a novel urban dataspace testbed and an illustrative use case. Among the key challenges are new categories of policy conflicts in urban data, complexity in multilateral data governance, and real-time sensor compliance scenarios. Key features are proposed to extend an existing Open Digital Rights Language (ODRL)-based policy engine to address these challenges, exploiting KG technologies and capabilities of large language models, including a concise formal semantics for ODRL, and enhancements to key components of the policy engine. We also provide concrete examples to illustrate the challenges and how this enhanced ODRL-based framework addresses them in the smart cities setting.

Keywords: Urban Dataspace · Knowledge Graph · ODRL · Data Governance · Policy Enforcement.

1 Introduction

The rapid expansion of interconnected sensors, systems and stakeholders is transforming smart cities into highly complex, data-intensive environments. Urban services now rely heavily on data-driven systems to manage resources, optimise operations, and improve citizens' quality of life. An urban dataspace is a shared environment where data flows among diverse actors and systems. In such context, urban data has become a critical asset, demanding appropriate governance to ensure its quality, security, privacy, and ethical use.

* Equal contribution.

Urban data is generated rapidly and continuously from heterogeneous sources, often in varying formats, granularities, quality, and update frequencies. Moreover, these data are typically shared and reused by multiple stakeholders with divergent objectives, responsibilities, and enforcement powers. This raises critical concerns related to multi-actor, dynamic and real-time policy enforcement, and the negotiation of usage rights and constraints across organisational boundaries.

The heterogeneity of data sources is a fundamental challenge in urban dataspace contexts. Knowledge Graph (KG) technologies can help address this by enabling semantic interoperability through shared vocabularies, flexible data models, and reasoning, thereby supporting accurate, trustworthy, and collaborative data sharing across domains [20]. In addition, policy enforcement—ensuring that data usage complies with regulations and expectations—represents another critical challenge. While policy frameworks such as the Open Digital Rights Language (ODRL) offer a foundation for expressing usage policies, current implementations fall short in supporting the emerging challenges specific to urban contexts, such as the spatio-temporal, IoT-aware, and dynamic policy enforcement. Existing tools lack adequate mechanisms to manage heterogeneous, fine-grained, context-aware constraints. There is also a lack of evidence-based research reporting the challenges and lessons learned from real-life urban dataspaces.

To explore the potential of applying KG technologies in urban contexts, this paper introduces the Urban Dataspace, initiated by GATE Institute – the International Data Space Association (IDSA) Hub in Bulgaria, and presents key policy challenges identified from real-world urban use cases. The challenges can be covered by ecosystem mechanisms or policy engines. While existing dataspace solutions, such as IDS-RAM [18], provide generic guidelines or mechanisms related to the ecosystem (roles, responsibilities, templates, etc.), they lack guidelines and adaptations to concrete use cases. In this paper, we focus on the challenges that can be addressed by policy engines, in particular, adapted to the specific requirements that emerged from the urban context, and we propose key extensions to an existing ODRL-based policy framework [32,14] in the smart city context. The extensions enable spatio-temporal and dynamic policy enforcement, covering crucial support related to data heterogeneity, fine-grained temporal conditions, actor dynamism, and sensor-context awareness.

The main contributions of this work are:

- A detailed analysis of the challenges for data governance regarding policy enforcement in urban dataspaces derived from real-life smart city use cases.
- A design proposal for extensions to an ODRL-based policy engine supporting spatio-temporal, multi-actor, and IoT-aware policy enforcement, and a semantic grounding that enables comparison between policies.

The rest of the paper is structured as follows: Section 2 describes background and related work. Section 3 presents a novel urban dataspace and policy challenges derived from associated urban use cases. Section 4 presents the ODRL policy engine extensions to address the identified challenges and an illustrative example. Finally, Section 5 concludes the paper and outlines future work.

2 Background and Related Work

2.1 Dataspace and Dataspace Protocol

A dataspace is an "interoperable framework, based on common governance principles, standards, practices and enabling services, that enables trusted data transactions between participants" [8]. A dataspace enables data sharing between various trusted participants, providing both a multi-organisational agreement and a supporting technical infrastructure. It answers the technology challenges of data sharing related to publishing and finding data, maintaining control over the shared data, negotiating the agreements between participants, etc. Sharing data between participants requires a metadata provision to facilitate the transfer of datasets through a data transfer protocol. The Dataspace Protocol [9] specifies how the metadata is delivered, defining how datasets are offered as Data Catalogue Vocabulary (DCAT) *catalogues* and how usage control is expressed as *Policies* with ODRL. It also prescribes how data usage *Agreements* are syntactically expressed and electronically negotiated, and how datasets are accessed based on Transfer Process Protocols. The primary purpose of the Dataspace Protocol is to support interoperability on several levels, like technical and semantic interoperability for participants, as well as interoperability on trust, organisational and legal levels. The Participants interact in the Dataspace through so-called Participant Agents, which are *Connectors*, implementing the Dataspace Protocol. The Connectors can implement additional internal functionalities, like monitoring or policy engines. The identity management is performed by the Identity Provider, which also validates additional claims. It is responsible for the provision of a Trust Framework and can be implemented as a centralised system, a decentralised system or a federated system.

DCAT enables a data provider to describe datasets and data services in a catalogue based on a standard model and vocabulary, enabling the consumption and aggregation of metadata from multiple catalogues and increasing the discoverability of datasets and data services. A decentralised approach to publishing data catalogues can be implemented, allowing federated search for datasets across catalogues in multiple locations. The interaction between a data provider and a data consumer starts with a *Contract Negotiation*, which is tracked through a series of states such as requested, offered, accepted, agreed, verified, finalised and terminated. After the contract agreement is accepted, the *Transfer Process* between the data provider and data consumer can be initiated.

2.2 Existing approaches for policy enforcement in urban contexts

Policy enforcement implements mechanisms, rules, and technologies that are put in place to ensure that data is used in accordance with established regulations and expectations [7]. This represents one of the significant challenges for urban dataspaces [4].

High-level approaches to policy enforcement in urban dataspaces typically span four interrelated dimensions: governance frameworks, technical mechanisms, regulatory compliance, and trust-building practices [15].

Organisational and data governance frameworks define organisational structure, processes, and compliance, e.g. defining clear policies and rules (e.g. access, usage, and consent) [7], legal/ethical foundations, approaches to multi-stakeholder governance models (e.g. centralised, decentralised or federated) [22].

Technical enforcement approaches automate elements of policy enforcement through access [24] and data usage [17] control mechanisms. Additional techniques include privacy-enhancing technologies (PETs) [1], blockchain-based (e.g., smart contracts), metadata governance and auditing/monitoring solutions.

Legal and regulatory compliance is essential to align enforcement with frameworks like the GDPR, Data Governance Act [11], and specific contractual agreements (e.g., data transaction or participation agreements).

Trust and accountability support legitimacy and transparency through auditability, independent oversight, ethical guidelines and engagement [23].

Prominent initiatives and frameworks providing foundations for orchestration of enforcement implementation include [21]: International Data Spaces Association (IDSA) [16], FIWARE Foundation [12], Gaia-X [13,25], Data Spaces Support Centre [6], BSI [3], iSHARE [19].

Open Digital Rights Language (ODRL) [26] is a flexible, extensible W3C standard that allows users to define machine-readable policies that may control usage of content and services. This is expressed through permissions (actions allowed), prohibitions (actions forbidden), duties/obligations (actions that must be performed) and constraints (conditions under which the former apply). Its application has expanded significantly within environments like urban dataspaces.

While ODRL describes policies, it doesn't inherently enforce them. Enforcement relies on data governance through a separate "policy execution engine" that interprets ODRL policies and translates them into actionable controls. Challenges and limitations exist, including precision of semantic definitions of how to process and evaluate policies [5], limited expressiveness for complex scenarios, lack of standardised implementation specification for operands and operators, limited capabilities for conflict resolution and integration complexities.

3 Urban Dataspace and Challenges

This section presents the Urban Dataspace, the first dataspace in Bulgaria, which is initiated and deployed by the Big Data for Smart Society (GATE) Institute.

3.1 Architecture and Components

The initial deployment of the Urban Dataspace is based on the IDS reference architecture model (IDS-RAM) [18], including the following building blocks in addition to the IDS Connector: Identity Provider, App Store, Metadata Broker, Clearing House and Vocabulary Hub.

With the release of the Dataspace Protocol, the Urban Dataspace has been migrated to the Eclipse Specification, which provides a comprehensive framework as a set of Eclipse Dataspace Components (EDC) with a basic set of features (functional and non-functional) for the implementation of dataspaces. It supports interoperability by design, leveraging the framework's defined APIs. The framework's components, shown in Figure 1, provide functionality that is mandatory for dataspaces as follows: (1) *Catalogue service*, publishing and securing assets that can be shared with other organisations; (2) *Identity service*, managing and verifying organisational credentials using, for example, OAuth2 tokens; (3) Control plane services, including the creation and processing of data usage agreements that grant access to data; and (4) *Data plane and monitoring services*, initiating and managing data transfers using the HTTP protocol.

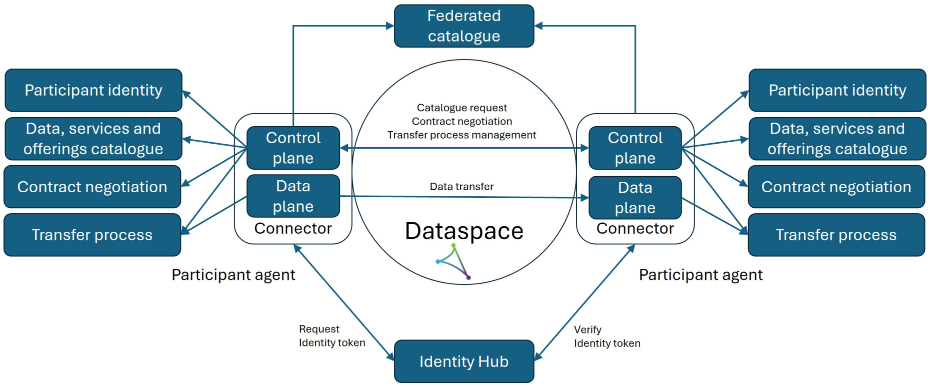

Fig. 1: Dataspace architecture and components.

The *Connector* consists of components for controlling data sharing and executing data transfer, namely the *Control Plane* and the *Data Plane*. The contract agreements that grant access to data, manage data transfers, and monitor usage policy compliance are managed by the Control Plane. The crawling and caching of data catalogues is performed by the *Federated Catalogue*. To determine whether it trusts and should grant data access to a counterparty, a participant connector uses verifiable credentials presented by the Identity Hub.

3.2 Real-world Policy Challenges in an Urban Dataspace

Urban Dataspace Context. Challenges in any dataspace, and particularly that we faced in an urban context, stem from the heterogeneity of data sources, providing data at varying spatial and temporal resolutions, the prevalence of incomplete or low-quality datasets, data sparsity, and the diversity of users with differing aims and roles. Data streams range from high-volume remote sensing sources, such as LiDAR scanning, Earth observation, and airborne sensing, producing data at rates of several gigabytes per minute, to low-volume, asynchronous sensors, including air quality and noise monitoring stations, operating at variable reporting intervals. Static spatial datasets from cadastral systems and diverse dynamic data flows, such as mobility feeds, parking occupancy, and pedestrian movement patterns, contribute to the overall information complexity. From the user perspective, typical urban stakeholders include data providers (e.g., municipal departments, utilities, private operators), data consumers (e.g., urban planners, researchers, startups), service providers and their clients (e.g., analytics vendors and transport operators), and governing authorities (e.g., municipal administrations, regulators). Although these actors operate within a shared technical infrastructure and regulatory framework, their objectives and priorities diverge. These differences influence access rights, permissible use contexts, temporal and spatial constraints, applicable obligations, and the manner in which rights and restrictions extend to derivative products. The deployment burden also differs in kind and weight across stakeholders. Not only do teams bring uneven legal, data, and IT literacy, but they also have different levels of process maturity, tooling, and capacity. For instance, stakeholders often face procurement and DPIA/security reviews that impose formal approvals and clearances (often including certification) which can force design changes, constrain data scope and contracts, and extend schedules. As data providers, they must translate licenses into approved templates, attach rich metadata to data products, and ensure policy inheritance for derived products. As data consumers, they must express intended use in machine-readable terms, while IT operations integrate connectors, identity, and audit logging with existing systems without disruption. These tasks are demanding and often block adoption due to limited staff, expertise, or tooling; however, many steps can be automated with the policy engine, proposed in Section 4, lowering barriers and enabling broader participation in urban dataspaces.

Use Case Development. Use cases are specific scenarios where participants are able to create economic, social, or environmental value through data sharing. The Use Case Development building block provides an operational environment for shaping new business models. This process unfolds in a series of elements:

1. Identification and Monitoring of Use Case Scenarios: Potential use case scenarios are brainstormed by leveraging various sources like participant needs, competitor analysis, and other data spaces to gather ideas.

2. Refinement of Use Case Scenarios: Promising use case scenarios are specified in detail, including the purpose, participants, data flows, and value propositions for each participant. The Data Cooperation Canvas, Use Case Playbook or self-created templates can support this refinement process.

3. Brokering and Ensuring Synergies: The data space connects the right participants and resources for use cases, fostering synergies between use cases. For example, multiple use cases can use common data products or infrastructure, reducing costs and boosting efficiency.

4. Support for Value Sharing in Use Cases: A critical element for successful use cases is to ensure the costs generated by it for the data space infrastructure are properly covered. Subscription fees or data exchange-based payments, for example, can facilitate this value distribution.

5. Ensuring Adherence to Principles: Use cases must also adhere to dataspace principles like data sovereignty and trust, along with any additional rules established at the dataspace or individual use case level.

6. Use Case Implementation: Once a use case scenario is refined, it is transformed into a functioning use case. This might involve further infrastructure development, data product creation, participant onboarding, agreements, including access and usage policies, and technical configurations.

7. Use Case Continuous Improvement: The performance of the use case is continuously analysed to identify possibilities for improvement. Changes are planned and implemented systematically throughout the life cycle of a use case.

Problems and Challenges. Consideration of the policy enforcement challenges from the use case development has led to the identification of problem areas (**Px**) and related policy challenges (**Cx**).

Heterogeneous Temporal and Spatial Granularity (P1): Datasets have different update frequency, causing *temporal mismatches* (e.g., sensors may report per minute, hour or day). Fine-grained policies may not apply consistently due to *spatial resolution differences* (e.g., point, street segment, neighbourhood, citywide). Hence, policy constraints may not map cleanly to what the data can support, risking under- or over-enforcement. **C1:** *Defining and enforcing consistent policies across uneven resolutions without losing critical context or introducing unfair bias.*

Schema and Semantic Heterogeneity (P2): Data providers describe the same phenomena with *different schemas, labels and units* (e.g. time may appear as “time”, “t”, or “timestamp”, measured in hours, seconds, or ISO 8601 strings) while data consumers and service providers express intended use in their own terms and baseline vocabularies. Policy rules that reference such concepts and units depend on correct alignment; when mismatches occur, checks on usage conditions or licensing constraints may yield false denials or unsafe permissions. **C2:** *Policies must be semantically aware to avoid misenforcement due to schema mismatches.* This requires normalisation or semantic mediation layers.

Data Reliability and Trustworthiness (P3): Datasets *differ widely in quality and origin* (e.g. sensory data may come from certified measurement stations or community-maintained devices), while key metadata is often incomplete, non-comparable, or entirely missing (e.g., calibration dates, accuracy/error bounds). Policy constraints often rely on such quality indicators (e.g., “allow access only to data with known error margins” or “train models only on recently calibrated sensors”), but in their absence or inconsistency, such policies may be misleading, overly restrictive, or silently bypassed. **C3:** *Enforcing policies which rely on data confidence is difficult without standardized reliability metrics.*

Policy Conflict and Ambiguity (P4): Datasets covering similar domains, such as traffic flows, air quality, or mobility, may come with *diverging access rules:* one provider allows open reuse, another enforces strict time-bounded conditions. Users or services consuming or aggregating such datasets face unclear or incompatible obligations, especially when the data is republished or combined.

This undermines legal certainty and hinders automation. **C4:** *Need mechanisms to detect, reconcile, or prioritise conflicting policies across sources.*

Dynamic and Context-Aware Requirements (P5): Certain policies are stateful and situational: clauses depend on geofences (e.g., in the EU only), time windows (rush hour only), and events (accidents, maintenance). Context may be specified by providers, yet consumers need timely decisions while data arrives as a mix of live streams, delayed batches, and backfills. It is difficult to determine consumer compliance as context shifts. **C5:** *Enforcing spatiotemporal and event-triggered policies in heterogeneous, sometimes delayed, data flows.*

Licensing and Usage Constraints (P6): Datasets arrive under *varied licensing models* (open, non-commercial, purpose-limited) with terms in heterogeneous templates or plain prose. Service outputs also carry policies that must be declared and inherited. Stakeholders (providers, consumers, service providers, governance) must author, compare, negotiate, and combine terms, decide access, and propagate rights and obligations to derivatives. When terms are non-standard or only human-readable, machines cannot reliably assess intended use, verify obligations, or detect cross-asset incompatibilities, leading to overblocking or non-compliant reuse. **C6:** *Automating the enforcement of legal and policy terms when licenses are non-standard or human-readable only.*

Multilateral Data Governance (P7): *Shared data assets* span public authorities, private operators, and citizen groups, each with its own governance model, accountability, and risk posture. As data and services cross boundaries, joint obligations, tiered access, and decision rights must be aligned to avoid conflicts and stalemates. **C7:** *Coordinating policy across stakeholders with divergent goals and enforcement powers.*

Data Ownership (P8): Europe *lacks a clear and universally accepted definition of data ownership*, and the way different existing ownership models should interact remains uncertain. The Data Act does not establish rules on data ownership itself. Instead, it centres on the concept of data holders, rather than data owners, and sets out regulations governing how third parties may access and use data in the possession of these holders, regardless of any actual or asserted ownership claims. **C8:** *Allocation and enforcement rights in an ecosystem where ownership concepts are ambiguous and potentially conflicting.*

Some of the challenges we faced in practice are inherently human-factor and outside any technical stack: ambiguous or shifting policies, inconsistent/missing metadata, mis-calibrated sensors, misaligned incentives, and legal review (i.e., issues needing governance/training/contracting rather than code). Where automation is viable, we separate ecosystem and engine duties. The ecosystem, mostly realised by the dataspace's control/identity planes, defines baseline rules and templates, negotiates exceptions, certifies connectors/providers, and handles dispute resolution and tiered access. The policy engine, proposed in Section 4) provides: (i) a formal layer for policy equality/containment/merge to detect conflicts and ensure stricter-than inheritance; (ii) an ontology/mediation layer (DCAT, SSN/SOSA, QUDT, PROV-O) to normalise schemas, units, and provenance; (iii) automated evaluation of credentials, provenance completeness, calibration/accuracy signals, and spatiotemporal clauses; and (iv) decisions and audit logs for runtime usage control via the connector.

3.3 Air Quality Use Case

The Urban Dataspace provides a secure digital infrastructure that enables trusted and sovereign data and service sharing among various stakeholders. It supports the implementation of the Urban Digital Twin (UDT) of Sofia city, covering use cases related to urban planning, air pollution and climate change mitigation, mobility and transportation, energy efficiency and assessment of building solar potential. To avoid complexity, this section presents the air quality use case as a sample scenario for showcasing the Urban Dataspace.

Dataspace participants. The air quality data is shared as Data Offers by four main data providers shown in Figure 2.

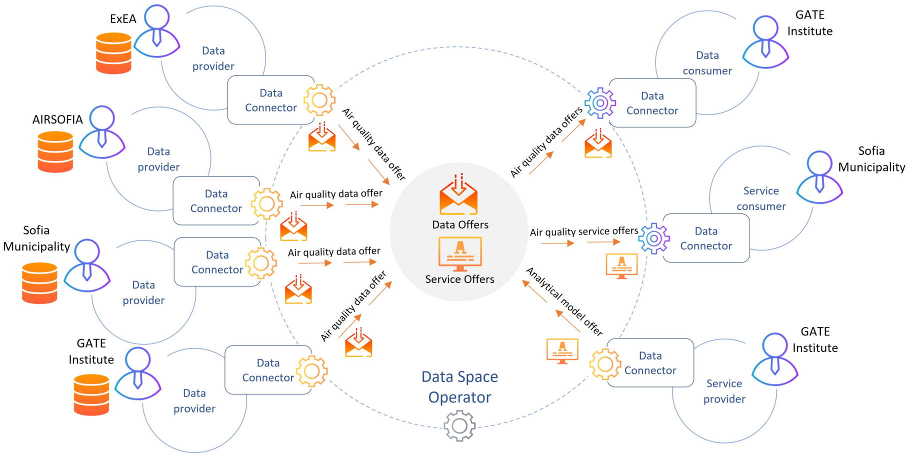

Fig. 2: Air quality scenario within the Urban Dataspace.

Bulgarian Executive Environment Agency (ExEA)[5] is an administration under the Minister of Environment and Water and a National Coordination Centre of the European Environment Agency. As such its data is considered the most reliable. *AIRSOFIA*[6], part of the Sensor.community's projects about Fine Particulate Matter (PM2.5) measurement in the environment. The data originate from low-cost community sensors, not subject to official calibration or regulatory standards. As a result, the measurements may exhibit higher uncertainty

[5] `https://eea.government.bg/en`
[6] `https://airsofia.info`

and variability compared to certified monitoring stations, limiting their reliability for precise air quality assessments or policy enforcement. *Sofia Municipality*[7]*,* collecting air quality data from 21 air quality stations that, while not officially certified like those of the ExEA, offer a higher degree of reliability compared to community-based sensors. These stations typically employ more robust hardware and maintenance protocols, providing more consistent and accurate measurements than low-cost, crowd-sourced alternatives. *GATE Institute*[8]*,* collecting air quality data through its City Living Lab, equipped with 12 air quality stations, similar to those of the municipality.

Given that participants in the dataspace may assume multiple roles, the GATE functions as both a data consumer and a service provider. Air quality data is utilised to develop predictive models for spatial and temporal prediction, employing a Gaussian auto-regressive approach and various machine learning models. The predictive models are offered as services within the dataspace and accessed by the Sofia Municipality to support evidence-based decision-making.

Problems and Challenges. The problems and challenges outlined for the Urban Dataspace in Section 3.2 can be illustrated through the air quality use case as follows:

Heterogeneous Temporal and Spatial Granularity (P1): Sensors report per-minute, hourly, or daily; locations vary city-wide, and some sensors sit near sensitive sites (e.g., school areas), where policies often require fine-grained geofences (e.g., "no access within 500 m of schools").

Schema and Semantic Heterogeneity (P2): Providers expose similar concepts with different schemas and units; PM2.5 might be labelled differently across data providers as pm25, PM_2_5, fine_particles. Policies referencing fields or units require mediation.

Data Reliability and Trust (P3): ExEA stations use certified, precise sensors with reliable uptime and rich metadata, but the raw feeds still contain gaps, unnormalised values, and artefacts, requiring substantial cleaning and validation before analysis; AIRSOFIA includes low-cost sensors, some out of calibration. Accuracy/calibration-dependent policies must account for this variance.

Policy Conflict and Ambiguity (P4): ExEA and Sofia Municipality data are open for reuse, whereas AIRSOFIA's terms are ambiguous; combining them creates uncertainty about downstream obligations.

Dynamic and Context-Aware Requirements (P5): Private providers can apply conditional access (e.g., "allow during pollution spikes"), requiring evaluation over time, place, and event state.

Licensing and Usage Constraints (P6): Where licenses exist, they are often human-readable only, preventing automated assessment and inheritance of obligations.

[7] https://platform.airthings-project.com

[8] https://citylab.gate-ai.eu

Multilateral Data Governance (P7): ExEA, AIRSOFIA, the Municipality, and GATE have distinct operational models, necessitating tiered access and clear responsibilities.

Data Ownership (P8): The problem of ownership arises from overlapping contributions and unclear rules on how rights and responsibilities are shared when datasets are integrated and shared as a common data product.

4 Proposal for a Policy Engine for Urban Dataspaces

In this section we present the proposal of an architecture of an enhanced policy engine and show how it addresses the urban dataspace challenges identified in Section 3.2 (referenced in text with code **(C1-C7)**). This proposal outlines key directions for advancing research and implementation efforts. Although no comprehensive solution currently exists for the challenges of urban dataspaces, several of the approaches reviewed in Section 2.2 could be adapted and integrated into our proposed framework. The architecture has been designed with scalability and usability in mind; however, a thorough evaluation of these dimensions—both theoretically and in practice—remains an important avenue for future work. Figure 3 illustrates the overall architecture, showing how its five key components align with the challenges.

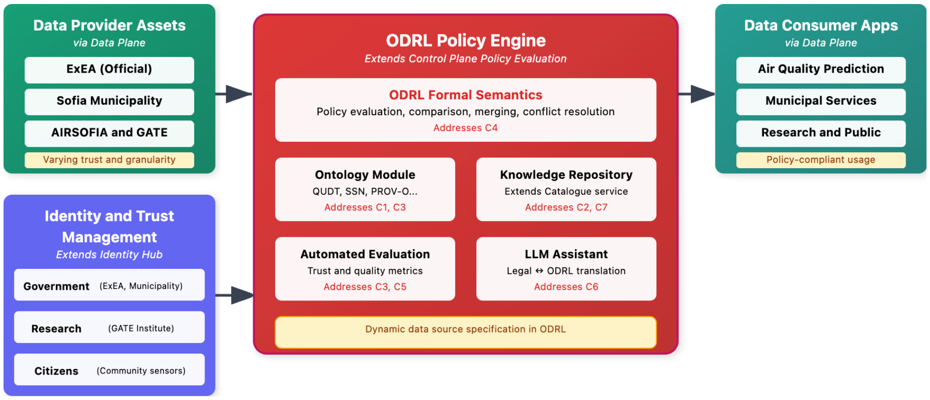

Fig. 3: ODRL-based Policy Engine Architecture for Urban Data Space

4.1 ODRL Formal Semantics (C4)

A formal semantics of ODRL is needed to provide a consistent approach to evaluate and compare policies, as required by challenge **(C4)**. However, such semantics is currently not available. While a number of partial formalisations exist, such as [10] and [2], these are focussed on particular features of ODRL, do not consider the latest ODRL version, namely version 2.2, or are preliminary works. To address the needs of an urban data-space, a suitable ODRL semantics should specify the semantics of multiple decision problems, such as (1) the evaluation of a policy on a state of the world, or on a data access request (e.g. does policy X allow me to download and share derivative work from dataset Y?), (2) the comparison of policies (e.g. are two policies semantically equivalent? Do the permissions of one imply the permissions of the other?) and (3) the merging of multiple policies to characterize the conditions of use of data derived from multiple sources. Moreover, a formal semantics should be simple, intuitive, easily implementable and computationally efficient. For example, in a data sharing scenario, a practical and cautious approach is to state explicitly what is permitted, rather than listing everything that is forbidden. Thus the chosen semantics can assert that everything not explicitly permitted is forbidden.

4.2 Ontology Module (C1, C3, C5)

Interoperability in an Urban Dataspace requires concepts and policies to be expressed in a mutually understandable way. While no single ontology covers all use cases, reuse of a core set is essential. In this regard, spatial and temporal resolution in policies can be expressed using DCAT properties such as `dcat:temporalResolution`, while sensor data and units of measurement are expressed with QUDT.[9] Custom vocabulary for common granularities (e.g., “street level”) enable greater expressivity **(C1)**. Data reliability depends on quality and provenance: Sensor, Observation, Sample, and Actuator Ontology (SOSA/SSN) models sensor accuracy, the Provenance Ontology (PROV-O) describes data derivation, and the Verifiable Credentials Data Model (VC) [31] provides certifiable claims, linking datasets to trusted issuers **(C3)**.

To regulate policies, ODRL constraints compare operands (e.g., PM2.5 values > 55.5 µg/m^3). However, specifying trusted data sources is critical. We extend ODRL with `leftOperandSource`, which points to specific providers, APIs, or models. Provenance and versioning are expressed with the Provenance, Authoring and Versioning (PAV) ontology [28], ensuring that when multiple sources exist, the chosen one is transparent and verifiable **(C5)**.

4.3 Knowledge Repository Module (C2, C7)

This module introduces repositories for both ontologies and policies. The ontology repository extends IDS-RAM’s Vocabulary Hub by not only storing vocabularies but also mappings across them, supporting RDFS [29] , SKOS [30] and OWL [27] terminology. These can specify when two concepts are the same, or when one is a more specific/general version of another. Adoption statistics will also promote convergence toward widely reused ontologies **(C2)**.

The policy repository stores ODRL agreements linked to regulated datasets, ensuring transparency of data usage. Stakeholders can review how datasets are

[9] `https://qudt.org`

being shared and their restrictions. For confidentiality, repositories allow selective visibility, e.g., granting access only to verified IDs. Redacted versions may be made public, while sensitive details remain private. This balance of openness and control supports accountability while protecting sensitive data **(C7)**.

4.4 Automated Evaluation Module (C3)

The Automated Evaluation Module provides scalable mechanisms to certify datasets and enrich them with quality and trust indicators. By generating and validating standardised metadata, it strengthens dataset discoverability and reliability across the dataspace **(C3)**. We envision this as a set of components, such as the following ones. The *ontology compliance* component can analyse datasets and their metadata to extract information about which ontologies it reuses. This information can be published alongside the dataset to enhance discoverability and provide quality measures related to its standardisation. The *granularity compliance* component analyses a dataset and, if it contains temporal and location data defined in a standard format, evaluates its temporal and/or spatial granularity. This information can then be used to annotate the dataset with this quality metric, or, if this quality metric already exists, to verify its correctness. The *digital certification* component is an extension of the Certificate Authority component, which allows for the verification of a wide array of facts, including dataset metadata, using the VC ontology. Together, these automate the evaluation of compliance, quality, and trust, reducing manual overhead.

4.5 LLM Assistant Module (C6)

The LLM Assistant connects machine-readable ODRL policies with human-readable legal text **(C6)**. Its functions are twofold: explaining ODRL policies in natural language for non-technical users, and converting between ODRL and legal text for expert review. For general users, it provides clear policy explanations and supports conversational queries (e.g., "Can I publish derived data if I remove location fields?"). For legal contexts, it proposes draft translations between ODRL clauses and legal formulations, though expert validation remains necessary due to the impossibility of guaranteeing correctness. This capability fosters transparency by enabling stakeholders to compare policies in different formats across the dataspace, while also highlighting potential conflicts or violations for further investigation. By serving as an interface between legal and technical domains, it ensures policies are both enforceable and understandable. Besides the task of policy translation, LLMs could also assist with secondary tasks, such as extracting ontological information from legal texts.

4.6 Example Scenario

We illustrate the policy engine with a practical case. Here, a policy permits use of air quality data only if the spatial resolution is up to 500 meters (e.g. district level) and the temporal resolution is hourly or finer. If these conditions are not met (e.g., faulty sensor output), the data may still be used, but only as background for trend analysis. The **Ontology Module** and **ODRL Formal Semantics** are shown in the following policy formalization. The first rule grants permission to use the dataset, while the second restricts usage to trend analysis when spatial or temporal limits are exceeded. Other actions (e.g., transfer of ownership) remain prohibited by default since they are not explicitly allowed.

```
ex:airQualityPolicy a odrl:Policy ;
 odrl:permission [
  odrl:target ex:AirQualityDataset ;
  odrl:action odrl:use ;
  ] ;
 odrl:prohibition [
  odrl:target ex:AirQualityDataset ;
  odrl:action odrl:use ;
  odrl:constraint [
   odrl:leftOperand dcat:spatialResolutionInMeters ;
   odrl:operator odrl:gt ;
   odrl:rightOperand "500"^^xsd:decimal
   ] ;
  odrl:constraint [
   odrl:leftOperand dcat:temporalResolution ;
   odrl:operator odrl:gt ;
   odrl:rightOperand "PT1H"^^xsd:duration
   ] ;
  odrl:constraint [
   odrl:leftOperand odrl:purpose ;
   odrl:operator odrl:neq ;
   odrl:rightOperand ex:backgroundTrendAnalysis
   ]
 ] .
```

Fig. 4: Example ODRL policy for the sharing of Air Quality Data.

When a consumer requests a dataset subset, the **Automated Evaluation Module** verifies spatial/temporal resolution and communicates usage permissions via the **LLM Assistant Module**. This ensures utility, maintains compliance, and provides graceful fallbacks in heterogeneous sensor networks, directly addressing the granularity mismatch problem (**C1**).

5 Conclusion and Future Work

In this paper, we identified eight problem areas and the associated key challenges for policy enforcement that emerged in urban dataspaces and the limitations in the existing ODRL solutions. To bridge this gap, we proposed the design of an enhanced policy engine grounded on a concise ODRL formal semantics for policy definition and validation, and the extensions to core modules in this context. Specifically, our proposal demonstrates how KG technologies (such as ontologies and knowledge repositories) and LLMs can be utilised to address the real-life urban challenges for policy enforcement and support for non-technical users.

While this paper focuses on urban policy challenges that can be addressed technically through policy engines, we also acknowledge human and organisational factors related challenges as highlighted in Section 3.2. For example, the challenge related to data ownership requires an agreed definition of data ownership and harmonised data ownership models, while multilateral data governance needs organisational coordination in addition to technical solutions.

For future work, we will extend the formalisation of ODRL semantics, implement the policy engine extensions and validate them in real-life use cases using the Urban Dataspace testbed. We also plan to integrate the policy engine with the Dataspace Connector to support the replication of the solution.

Acknowledgments. The work is funded through the projects DataPACT (HE 101189771), DataTwin (BG16RFPR002-1.009-0005), CauseFinder (PNRR 760049), and UPCAST (HE 101093216).

Disclosure of Interests. The authors have no competing interests.

References

1. Auñón, J., Hurtado-Ramírez, D., Porras-Díaz, L., Irigoyen-Peña, B., Rahmian, S., Al-Khazraji, Y., Soler-Garrido, J., Kotsev, A.: Evaluation and utilisation of privacy enhancing technologies—a data spaces perspective. Data in Brief **55**, 110560 (2024)
2. Bonatti, P.A., Fornara, N., Harth, A.: Towards a formal semantics of the open digital rights language (odrl 2.2). In: Proceedings of the OPAL 2025 (2025)
3. BSI: Smart cities framework for sharing data and information services - pas 183. https://www.bsigroup.com/en-GB/insights-and-media/insights/brochures/pas-183-smart-cities-framework-for-sharing-data-and-information-services, accessed: 2025-08-18
4. Choenni, S., Bargh, M.S., Busker, T., Netten, N.: Data governance in smart cities: Challenges and solution directions. Journal of Smart Cities and Society **1**(1), 31–51 (2022)
5. Cimmino, A., Cano-Benito, J., García-Castro, R.: Open digital rights enforcement framework (odre): From descriptive to enforceable policies. Computers & Security **150**, 104282 (2025)
6. Data Space Support Centre: Data space support centre. https://dssc.eu, accessed: 2025-08-18
7. Data Spaces Support Centre: Access & usage policies enforcement. https://dssc.eu/space/BVE/357075567/Access+&+Usage+Policies+Enforcement, accessed: 2025-08-18
8. Data Spaces Support Centre: Data spaces blueprint v2.0. https://dssc.eu/space/BVE2/1071251781/1+Key+Concept+Definitions, accessed: 2025-08-30
9. Data Spaces Support Centre: Dataspace protocol. https://docs.internationaldataspaces.org/ids-knowledgebase/dataspace-protocolhttps://dssc.eu/space/BVE/357075567/Access+&+Usage+Policies+Enforcement, accessed: 2025-08-18
10. De Vos, M., Kirrane, S., Padget, J., Satoh, K.: Odrl policy modelling and compliance checking. In: International Joint Conference on Rules and Reasoning. pp. 36–51. Springer (2019)
11. European Commission: Data act explained. https://digital-strategy.ec.europa.eu/en/factpages/data-act-explained#:~:text=In%20addition%2C%20it%20introduces%20safeguards,the%20European%20economy%20and%20society, accessed: 2025-08-18
12. FIWARE: Fiware. https://www.fiware.org, accessed: 2025-08-18
13. Gaia-X: Gaia-x. https://gaia-x.eu, accessed: 2025-08-18
14. Gheisari, S., Yumusak, S., Salas, J.O., Ibáñez, L.D., Konstantinidis, G., Roman, D.: Towards modular data marketplaces. In: Proceedings of the 8th International Joint Conference on Rules and Reasoning (2024)
15. International Data Spaces: Guiding principles. https://docs.internationaldataspaces.org/ids-knowledgebase/idsa-rulebook/idsa-rulebook/2.-guiding-principles, accessed: 2025-08-18

16. International Data Spaces: International data spaces. https://internationaldataspaces.org, accessed: 2025-08-18
17. International Data Spaces: Usage control in the international data spaces. https://internationaldataspaces.org/wp-content/uploads/dlm_uploads/IDSA-Position-Paper-Usage-Control-in-the-IDS-V3..pdf, accessed: 2025-08-18
18. International Data Spaces Association: Ids-ram 4.0. https://github.com/International-Data-Spaces-Association/IDS-RAM_4_0, accessed: 2025-08-18
19. iSHARE: Trust framework for data rights. https://ishare.eu, accessed: 2025-08-18
20. Lam, A.N., Avogadro, R., Martin-Recuerda, F., Elvesæter, B., Ma, X., Nystad, E.J., Roman, D., Berre, A.J.: Towards a toolkit for semantic interoperability in data spaces. In: 2025 IEEE 8th International Conference on Industrial Cyber-Physical Systems (ICPS). pp. 1–7 (2025). https://doi.org/10.1109/ICPS65515.2025.11087848
21. Martella, A., Martella, C., Longo, A.: Designing data spaces: Navigating the european initiatives along technical specifications. arXiv:2503.1599 (2025)
22. Micheli, M., Ponti, M., Craglia, M., Berti Suman, A.: Emerging models of data governance in the age of datafication. Big Data & Society **7**(2) (2020)
23. Perucci, F., Swanson, E.: Building trust and facilitating use of data. Statistical Journal of the IAOS **40**(1), 71–79 (2024)
24. ProdataKey: Understanding access control systems: What is access control? https://www.prodatakey.com/single-post/what-is-an-access-control-system, accessed: 2025-08-18
25. Siska, V., Karagiannis, V., Drobics, M.: Building a dataspace: Technical overview. https://www.gaia-x.at/wp-content/uploads/2023/04/WhitepaperGaiaX.pdf, accessed: 2025-08-18
26. W3C: Ordl information model. https://www.w3.org/TR/odrl-model, accessed: 2025-08-18
27. W3C: Owl web ontology language. https://www.w3.org/TR/owl-features, accessed: 2025-08-18
28. W3C: Pav - provenance, authoring and versioning. https://pav-ontology.github.io/pav, accessed: 2025-08-18
29. W3C: Rdf schema 1. https://www.w3.org/TR/rdf-schema, accessed: 2025-08-18
30. W3C: Skos simple knowledge organization system namespace document. https://www.w3.org/2009/08/skos-reference/skos.html, accessed: 2025-08-18
31. W3C: Verifiable credentials data model v2.0. https://www.w3.org/TR/vc-data-model-2.0, accessed: 2025-08-18
32. Yumusak, S., Gheisari, S., Salas, J.O., Ibáñez, L.D., Konstantinidis, G.: Data sharing negotiation and contracting. In: Proceedings of ISWC 2024 (2024)