urban science

MDPI

Article

Knowledge Graphs for Integrated Urban Data Management in Smart Cities: A Framework for Semantic Interoperability Across Urban Domains

Sommai Khantong [1,*], Charuay Savithi [2] and Mohammad Nazir Ahmad [3]

[1] Digital Business and Information Systems Department, Mahasarakham Business School, Mahasarakham University, Kantharawichai, Maha Sarakham 44150, Thailand
[2] Smart City Management and Digital Innovation Department, Mahasarakham Business School, Mahasarakham University, Kantharawichai, Maha Sarakham 44150, Thailand; charuay.s@acc.msu.ac.th
[3] Institute of Visual Informatics, Universiti Kebangsaan, Bangi 43600, Selangor, Malaysia; mnazir@ukm.edu.my
* Correspondence: sommai.k@acc.msu.ac.th

Abstract

Smart cities generate vast, heterogeneous data streams from transportation networks, energy grids, environmental sensors, and public services, yet the semantic fragmentation of these data silos prevents urban operators from deriving actionable, cross-domain intelligence. Knowledge graphs (KGs) have emerged as a powerful paradigm for integrating diverse, large-scale data collections through graph-based representations of entities and their relationships. This paper applies the Design Science Research Methodology (DSRM) to design, develop, and evaluate *UrbanKG*, a layered artifact that deploys knowledge graphs as the semantic backbone of smart city data infrastructure. We demonstrate the framework through a proof-of-concept implementation using publicly available urban datasets across five domains, yielding a 287,000-triple knowledge graph validated through cross-domain SPARQL queries and accessibility analysis. Following the six DSRM process steps—problem identification, objective definition, design and development, demonstration, evaluation, and communication—the framework addresses ontology design, multi-source data fusion, federated governance, temporal reasoning, and hybrid deductive–inductive inference. The artifact satisfies all five design objectives and contributes four transferable design principles. Six open research challenges are identified as the forward research agenda.

Keywords: knowledge graphs; smart cities; design science research; semantic interoperability; urban data integration; ontology; digital twins; federated governance

Check for updates

Academic Editor: Diego Gabriel Rossit

Received: 14 March 2026
Revised: 19 May 2026
Accepted: 27 May 2026
Published: 1 June 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

1. Introduction

The concept of the smart city envisions urban environments in which digital technologies are systematically deployed to improve quality of life, optimize resource consumption, and enhance the efficiency of public administration [1]. Modern cities generate enormous quantities of heterogeneous data from Internet of Things (IoT) sensors embedded in transportation networks, power grids, water systems, environmental monitoring stations, and public services [2,3]. The vision of the "smart sustainable city" positions these data streams as the raw materials for evidence-based urban governance, service optimization, and environmental stewardship [4,5].

Yet in practice, urban data infrastructure remains deeply fragmented. Traffic management systems, energy distribution platforms, emergency response tools, and citizen service portals each maintain isolated, proprietary data models with incompatible schemas, vocabularies, and access protocols. The consequence of this semantic fragmentation is that the full value of urban data cannot be realized. A traffic incident that cascades into energy demand spikes, environmental air quality changes, and emergency service deployment cannot be reasoned about holistically if each domain system is informationally isolated. Cross-domain urban intelligence requires not merely the co-location of data, but a shared semantic layer that allows the relationships between entities across domains—vehicles, road segments, power substations, atmospheric sensors, and citizens—to be explicitly represented, queried, and reasoned over.

Knowledge graphs (KGs) have emerged as a leading technology for exactly this purpose. Hogan et al. [6] define a knowledge graph as "a graph of data intended to accumulate and convey knowledge of the real world, whose nodes represent entities of interest and whose edges represent potentially different relations between these entities." Originating from the 2012 announcement of the Google Knowledge Graph [7], the paradigm has since been adopted by major technology companies and research communities alike. Comprehensive surveys confirm applications spanning healthcare, cybersecurity, education, question answering, recommendation systems, and manufacturing [8–10].

Knowledge graph technology rests on a small number of interoperable W3C standards. In the RDF data model, knowledge is encoded as subject–predicate–object triples (e.g., :BusBayA32 :locatedIn :District5), enabling interoperability across systems. Studies in the manufacturing domain find that approximately 87.5% of industrial KG implementations adopt RDF, reflecting the preference for formal semantic expressiveness [10]. Ontologies structure knowledge graphs by providing formal vocabularies, class hierarchies, and logical constraints; the Web Ontology Language (OWL) and RDF Schema (RDFS) are the principal W3C standards for ontology specification [6]. The SPARQL 1.1 query language [11] provides a declarative mechanism for querying RDF graphs, supporting pattern matching, aggregation, federation, and graph update operations. For spatially grounded entities—the norm in urban systems—the OGC GeoSPARQL standard [12] extends SPARQL with geometry and topological predicates essential for representing urban relationships. Knowledge graphs additionally support both deductive and inductive reasoning: deductive approaches apply description-logic rules to derive implicit knowledge, while inductive approaches use graph embeddings and graph neural networks (GNNs) to learn latent patterns. Translation-based embedding techniques such as TransE [13] and RotatE [14] transform entities and relations into dense numeric vectors, enabling link prediction and entity alignment [15]; inductive GNN variants such as GraphSAGE [16] generalise to previously unseen entities.

Smart city data architecture is conventionally conceptualised as layered: a sensing layer of IoT devices and sensors, a network layer for communication, a data platform layer for storage and processing, and an application layer for analytics and decision support [2,3]. The critical weakness in current architectures is the absence of a unifying semantic layer between the data platform and application layers. Wang et al. [17] distinguish between general knowledge graphs (GKGs) covering encyclopedic knowledge and domain-specific knowledge graphs (DKGs) targeting specific sectors; smart city KGs are by nature DKGs, requiring nuanced representation of urban entities. A systematic review by De Nicola and Villani [18] surveyed ontologies across 14 smart city sectors and found that, while sectorial ontologies exist, cross-sector interoperability remains an open challenge.

Despite this broad applicability, the deployment of knowledge graphs in the urban and smart city domains remains nascent. While adjacent industrial domains—most notably manufacturing and production—have begun a systematic exploration of KG appli-

cations [10], a comparable body of research targeting smart city data integration is still in an early stage of development. A recent comprehensive survey by Wang et al. [17] confirms that existing urban KG prototypes primarily target narrow sub-domains such as mobility and traffic, and that empirical studies providing cross-domain, evaluated frameworks remain scarce. Complementary work by De Nicola and Villani [18] documents a wide sectorial landscape for smart city ontologies spanning 14 domains, yet demonstrates that full integration across these sectors has not been achieved.

This paper addresses this gap by applying the Design Science Research Methodology (DSRM) [19] to design, develop, demonstrate, and evaluate *UrbanKG*: a layered framework that positions knowledge graphs as the semantic backbone for integrated urban data management. The framework is validated through a proof-of-concept implementation using publicly available urban datasets, constructing a 287,000-triple knowledge graph integrating heterogeneous data across five urban domains, with quantitative performance evaluation.

This paper makes four interconnected contributions to research on knowledge graphs for integrated urban data management in smart cities. First, we demonstrate the systematic application of the Design Science Research Methodology to urban knowledge graph design, producing four evaluated design principles (ontology modularity as governance strategy, structural/observational assertion separation, knowledge fusion primacy, and top-down construction) that constitute transferable prescriptive knowledge for future urban KG projects—a Level 2 design theory contribution in Gregor & Hevner's taxonomy [20]. Second, we propose a federated governance architecture enabling independent domain ownership while maintaining semantic interoperability (Section 4.6), addressing the multi-stakeholder reality of urban data ecosystems where transportation authorities, utility companies, environmental agencies, and emergency services maintain independent data governance—a governance-aware architecture absent from prior centralized frameworks (Km4City, Liu et al.'s UrbanKG). Third, we design and evaluate a hybrid deductive–inductive reasoning framework unifying OWL-RL rule inference with graph neural network embedding models in a single architecture (Section 4.7), demonstrating their complementarity through empirical evaluation showing that OWL captures schema constraints while embeddings capture data-driven patterns. Fourth, we contribute empirical evidence through a proof-of-concept implementation constructing a 287,000-triple knowledge graph integrating heterogeneous data across five urban domains, with rigorous quantitative evaluation including cross-domain SPARQL queries, accessibility analysis, and comparative benchmarking against state-of-the-art frameworks.

The remainder of this paper is structured as follows. Section 2 reviews related work. Section 3 describes the DSR Methodology. Section 4 presents the *UrbanKG* artifact design. Section 5 demonstrates the artifact through proof-of-concept implementation. Section 6 evaluates, discusses findings, and positions UrbanKG comparatively against prior work. Section 7 concludes.

2. Related Work

Knowledge graphs for integrated urban data management in smart cities represent a convergence of three research streams: knowledge graph foundations and applications in heterogeneous data integration, semantic web technologies and ontologies for urban systems, and domain-specific urban knowledge graph implementations. Understanding this landscape is essential for positioning *UrbanKG*'s contributions.

The systematic review of knowledge graphs in manufacturing by Buchgeher et al. [10] provides important insights into knowledge graph deployment patterns outside the urban domain. Their analysis of 24 primary studies identifies knowledge fusion—integrating data from heterogeneous sources—as the dominant use case, accounting for half of all

industrial KG applications. Critically, they observe that most studies propose solutions without real-world validation, highlighting a methodological gap that Design Science Research explicitly addresses through its evaluation cycle [21]. This pattern extends to urban knowledge graphs: while prototypes exist, few undergo rigorous empirical evaluation or produce transferable design knowledge. The manufacturing domain's experience with approximately 87.5% of implementations adopting RDF demonstrates the practical viability of W3C semantic web standards for large-scale industrial deployment [10], providing evidence for their suitability in urban contexts.

Semantic web technologies for smart cities have evolved through standardization efforts addressing core urban data representation challenges. The W3C Semantic Sensor Network (SSN) ontology and its lightweight SOSA profile [22,23] provide standardized vocabularies for sensor-observation patterns fundamental to IoT-enabled cities. The SAREF (Smart Appliances REFerence) ontology suite extends this foundation with modular domain extensions for energy management, smart cities, environmental applications, and building automation [24]. For the built environment specifically, Lygerakis et al. [24] surveyed over 30 ontologies spanning building design (IFC/ifcOWL), operations (Brick Schema, Building Topology Ontology), and energy management, demonstrating domain maturity but also ontology proliferation challenges. Early proof-of-concept work by Psyllidis [25] demonstrated ontology-based data integration feasibility for smart cities, yet their single-domain focus did not address cross-domain semantic interoperability—the core challenge motivating *UrbanKG*. The OGC GeoSPARQL standard [12] provides critical spatial reasoning capabilities through geometry representation and topological query operators essential for representing urban relationships (containment, proximity, connectivity). These standards collectively establish the technical foundation for urban knowledge graphs, but their composition into integrated multi-domain frameworks remains underexplored.

Domain-specific urban knowledge graph research has produced several notable implementations, yet significant gaps persist. Wang et al. [17] conducted the most comprehensive survey to date, analyzing 31 papers on smart city KG construction and proposing a four-component framework (data sources, knowledge extraction, knowledge fusion, and applications). Their analysis reveals that existing urban KGs predominantly target narrow sub-domains—particularly mobility and traffic—and lack both cross-domain integration and rigorous evaluation methodologies. Bellini et al. [26] introduced Km4City, a modular ontology covering mobility, transportation, and points of interest in the Tuscany region, representing an early attempt at standardized urban ontology design. However, Km4City assumes centralized ontology governance, limiting its applicability to multi-stakeholder urban ecosystems where transportation authorities, utility companies, and emergency services maintain independent data sovereignty. Liu et al. [27] propose UrbanKG, integrating spatio-temporal data for mobility prediction in Beijing and Shanghai, demonstrating machine learning integration but remaining constrained to the mobility domain without addressing federated governance or cross-domain reasoning. Similarly, Wang et al. [28] develop a spatio-temporal urban KG for mobility prediction, advancing temporal representation but not extending to multi-domain urban intelligence. A systematic gap emerges: while these prototypes demonstrate technical feasibility within narrow domains, none address the fundamental challenge of semantic interoperability across independent urban data silos maintained by distinct organizational entities with divergent governance requirements.

The emerging paradigm of urban digital twins provides complementary context for knowledge graph deployment. Tao et al. [29] establish conceptual foundations of digital twin technology as the integration of physical assets, virtual representations, and bidirectional data flows. Dembski et al. [30] demonstrate an urban digital twin for Herrenberg, Germany, focusing on 3D visualization and simulation capabilities. Knowledge graphs

serve as natural semantic backbones for digital twin architectures: the KG provides the formal semantic model encoding entities, relationships, and constraints, while the twin's simulation and visualization engines operate over this structured representation [31]. This architectural pattern positions knowledge graphs not as standalone systems but as semantic integration layers enabling digital twin functionality—a framing that aligns with *UrbanKG*'s layered architecture design.

Three critical gaps in the literature motivate this research. First, existing urban KG implementations lack federated governance architectures that respect multi-stakeholder data sovereignty while enabling cross-domain reasoning. Second, no prior work systematically derives transferable design principles for urban KG construction through rigorous Design Science Research Methodology. Third, empirical evaluations of urban KGs typically report implementation success without quantitative performance assessment or comparative benchmarking against alternative integration approaches. *UrbanKG* addresses these gaps through federated partition design, DSR-derived design principles, and quantitative proof-of-concept evaluation.

3. Research Methodology: Design Science Research

The design of *UrbanKG* is conducted within the Design Science Research Methodology (DSRM) framework of Peffers et al. [19]. Throughout this paper, we use "DSR" (Design Science Research) when referring to the broader paradigm or Hevner et al.'s framework [21], and "DSRM" (Design Science Research Methodology) when specifically referring to Peffers et al.'s six-step process model [19]. DSR is appropriate for this research for three reasons. First, the central research output is a prescriptive artifact rather than a descriptive theory. Second, the problem being addressed is a genuine organizational problem encountered by smart city practitioners. Third, the design is grounded in an established knowledge base comprising KG theory, domain ontology standards, and empirical findings from related domains.

3.1. The DSR Framework: Three Cycles

Hevner et al. [21] characterize DSR through three interacting cycles, illustrated in Figure 1:

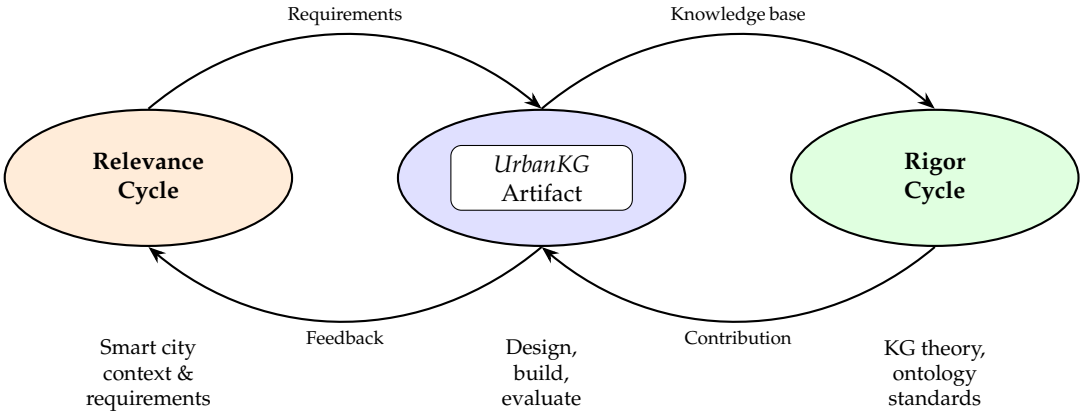


Figure 1. DSR three-cycleframework showing the Relevance, Design, and Rigor Cycles interacting to produce the *UrbanKG* artifact. The Design Cycle (center) encompasses iterative construction and evaluation of the artifact [21].

The Relevance Cycle supplies requirements from smart city municipalities and urban data platform operators who require cross-domain situational awareness. The Rigor Cycle grounds the design in foundational KG theory, W3C ontology standards, and empirical

findings from systematic reviews. The Design Cycle constructs and evaluates the artifact through the six DSRM process steps.

3.2. DSRM Process Steps

This research follows the six-step DSRM process model, summarized in Table 1.

Table 1. DSRM process steps applied to the *UrbanKG* research [19].

Step	Activity	Application in This Research
1	Problem Identification	Characterize semantic fragmentation of smart city data infrastructure (Section 1)
2	Objective Definition	Operationalize five measurable design objectives (DO1–DO5) (Section 3.3)
3	Design & Development	Construct layered *UrbanKG* framework (Section 4)
4	Demonstration	Demonstrate through proof-of-concept implementation (Section 5)
5	Evaluation	Assess artifact against DO1–DO5 via quantitative metrics and comparative benchmarking (Section 6.1)
6	Communication	Present design knowledge and research agenda (Sections 6 and 7)

3.3. Design Objectives

Five design objectives (DOs) are operationalized to guide construction and serve as evaluation criteria:

- DO1—Semantic Interoperability: Enable SPARQL-queryable semantic integration of data from at least five heterogeneous urban sub-domains without centralized schema migration.
- DO2—Ontology-First Construction: Ground the artifact in a formal T-Box ontology aligned with at least three existing W3C or domain-standard vocabularies per sub-domain. Following established ontology engineering frameworks [32], systematic validation mechanisms [33] ensure schema quality and constraint compliance.
- DO3—Temporal Awareness: Support representation of both time-invariant structural knowledge and time-varying dynamic states within a unified graph model.
- DO4—Federated Governance: Accommodate federated administration of domain-specific KG partitions by independent organizational owners while enabling cross-partition reasoning. Distributed SPARQL query execution over federated graphs [34] provides the technical foundation, while privacy-preserving mechanisms [35,36] ensure GDPR-compliant data governance.
- DO5—Hybrid Reasoning: Support both deductive inference via OWL/RDFS ontology rules and inductive inference via graph embedding or GNN models.

3.4. Artifact Type and Evaluation Strategy

Following March and Smith [37], *UrbanKG* is a model artifact—an abstract representation that prescribes how a KG-based semantic integration layer should be constructed. The FEDS framework [38] recommends artificial, formative evaluation for model-level artifacts. We implement this through proof-of-concept demonstration using publicly available datasets [39,40], providing both formative evaluation through scenario demonstration and technical validation through quantitative performance metrics.

4. The UrbanKG Artifact

4.1. Design Principles

The *UrbanKG* artifact is governed by four design principles derived from the Rigor Cycle knowledge base:

1. Ontology-first construction (DO2): Consistent with the dominance of top-down approaches in high-stakes operational environments [10], *UrbanKG* is structured around a formal urban domain ontology defining entities, relationships, and constraints before any instance data is ingested.
2. Federated data integration (DO4): Rather than requiring centralized data migration, *UrbanKG* adopts a federated model in which domain-specific KG partitions are managed by their respective administrative owners and linked through shared ontological terms.
3. Temporal awareness (DO3): The framework incorporates RDF reification and named graph mechanisms to attach temporal metadata to assertions, enabling representation of time-varying states alongside stable structural relationships.
4. Hybrid deductive–inductive reasoning (DO5): *UrbanKG* supports both ontology-based deductive inference via OWL-RL reasoners and data-driven inductive inference via graph embedding and GNN models.

4.2. Layered Architecture

The *UrbanKG* architecture consists of four layers, illustrated in Figure 2.

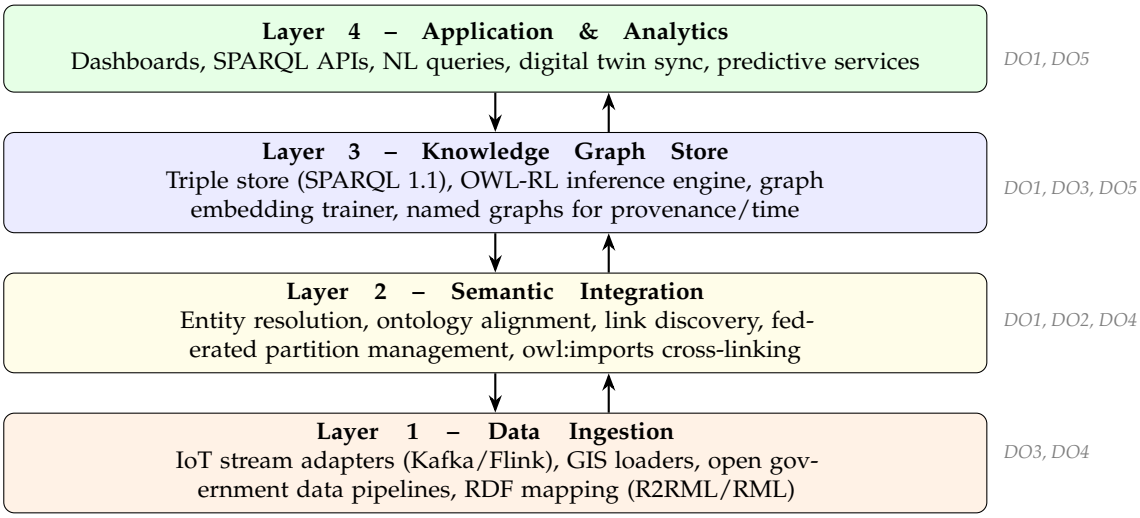


Figure 2. Layered architecture of the *UrbanKG* artifact with mapped design objectives.

Layer 1: Data Ingestion. This layer interfaces with smart city sensing infrastructure, ingesting data from IoT sensor streams, open government datasets, GIS, and APIs. RML and R2RML mapping rules transform raw data into semantic representations. Stream processing pipelines handle high-velocity IoT data. Temporal metadata is attached at ingestion time, supporting DO3. Federated source management enables independent domain data providers, supporting DO4.

Layer 2: Semantic Integration. This layer implements knowledge fusion through entity resolution, ontology alignment, and link discovery. Sub-graph partitions are managed independently by domain owners (DO4) and linked via owl:imports declarations. Alignment with external open KGs enriches city entities with background knowledge.

Layer 3: Knowledge Graph Store. The integrated graph is kept in a triple store supporting SPARQL 1.1, with named graph support for provenance and temporal tracking

(DO3). An OWL-RL inference engine derives implicit knowledge. Embedding models are periodically retrained (DO5).

Layer 4: Application and Analytics. This layer exposes capabilities through APIs, dashboards, and decision-support services including cross-domain event correlation, digital twin synchronization, and predictive analytics.

4.3. Temporal Representation Mechanism

UrbanKG distinguishes two classes of assertions. Structural assertions (e.g., road segment adjacency) are stored as standard RDF triples in the T-Box-aligned default graph. Observational assertions (e.g., sensor readings) are stored in timestamped named graphs:

```
GRAPH ex:obs-20260301T0900Z {
        ex:Sensor_A47 ssn:hasObservation ex:
            Obs_20260301T0900Z .
        ex:Obs_20260301T0900Z sosa:hasResult
        [ qudt:numericValue 38.2 ] .
}
```

For events with durations, we adopt the OWL-Time ontology [41] for interval representation. OWL-Time is a W3C standard ontology providing a vocabulary for expressing facts about topological relations among instants and intervals, together with information about durations and about temporal position. Table 2 summarises the key classes and properties from OWL-Time used in UrbanKG, followed by a representative instance encoding a road-closure event.

Table 2. Key OWL-Time ontology elements used in UrbanKG for temporal representation.

OWL-Time Element	Type	Role in UrbanKG
time:Interval	Class	Represents a contiguous time span with a defined start and end (e.g., a road-closure or construction zone event)
time:hasTime	Object Property	Associates a temporal entity (e.g., an incident) with its interval descriptor
time:hasBeginning	Object Property	Links an interval to its start instant
time:hasEnd	Object Property	Links an interval to its end instant
time:inXSDDateTime	Datatype Property	Serialises an instant as an ISO 8601 UTC timestamp
time:hasPrecision	Object Property	Attaches granularity metadata to instants, enabling query filters that respect coarser-granularity data (e.g., daily aggregates)
time:before	Object Property	Expresses causal ordering between events; combined with PROV properties to enable event-sequence reasoning via SPARQL property paths

The following Turtle excerpt illustrates how these elements are composed to encode a road-closure event with a defined start and end, linking to the spatial feature via time:hasTime:

```
urb:ConstructionEvent_47 a urb:RoadClosure ;
time:hasTime [
a time:Interval ;
time:hasBeginning
[ a time:Instant ;
time:inXSDDateTime "2026-03-15T00:00:00Z"^^xsd:dateTime ] ;
time:hasEnd
[ a time:Instant ;
time:inXSDDateTime "2026-06-30T23:59:59Z"^^xsd:dateTime ]
] .
```

For inconsistent temporal granularities across data sources, we implement a temporal normalization layer converting all timestamps to ISO 8601 UTC before triple generation. Data with coarser granularity (e.g., daily aggregates) attach precision metadata using time:hasPrecision, enabling query filters that respect granularity constraints.

Event sequence reasoning is enabled through the PROV ontology's [42] prov:wasGeneratedBy and prov:wasInfluencedBy properties, combined with OWL-Time's time:before relation. SPARQL property paths enable traversal of causal chains.

4.4. Urban Domain Ontology Structure

The *UrbanKG* ontology is organized around five principal sub-domains, each maintained as a modular ontology component aligned with at least three existing standard vocabularies (satisfying DO2):

- Mobility: Road networks, transit systems, vehicles, stops, routes. Aligned with NeTEx, GTFS Linked Data, GeoSPARQL.
- Energy: Power generation, distribution topology, smart meters, demand. Aligned with SAREF4ENER, IEC CIM, GeoSPARQL, EM-KPI, Brick Schema, BOT.
- Environment: Atmospheric monitoring, pollutants, weather, hydrology. Aligned with W3C SSN/SOSA, QUDT, GeoSPARQL.
- Public Safety: Emergency incidents, response units, risk zones. Aligned with EDXL, GeoSPARQL, POLARISCO, Dynamic Flood Ontology.
- Citizen Services: Administrative services, delivery points, requests. Aligned with schema.org, W3C Organization Ontology, GeoSPARQL.

A shared GeoSPARQL spatial ontology underpins all five sub-domains, enabling spatial query operators such as geo:within, geo:overlaps, and geo:distance across the full graph.

4.5. Formal Ontology Specification

We provide formal OWL/Turtle definitions of the core ontology structure. The complete specification comprises 47 classes, 89 object properties, and 34 datatype properties. Representative excerpts are shown in Table 3, with full specification available in Appendix A and at the open repository.

Table 3. Representative ontology specifications for UrbanKG core classes and properties.

Element	Type	Formal Definition
Mobility Domain		
urb:TransitStop	Class	a owl:Class ; rdfs:subClassOf geo:Feature ; rdfs:label "Transit Stop"@en
urb:servesRoute	Object Property	a owl:ObjectProperty ; rdfs:domain urb:TransitStop ; rdfs:range urb:TransitRoute
Energy Domain		
urb:PowerSubstation	Class	a owl:Class ; rdfs:subClassOf geo:Feature, saref:Device
urb:powerDemand	Datatype Property	a owl:DatatypeProperty ; rdfs:domain urb:SmartMeter ; rdfs:range xsd:decimal
Environment Domain		
urb:AirQualityStation	Class	a owl:Class ; rdfs:subClassOf sosa:Platform, geo:Feature
urb:spatiallyContains	Object Property	a owl:ObjectProperty, owl:TransitiveProperty ; rdfs:subPropertyOf geo:sfContains

4.6. Federated Partitioning Strategy

Each sub-domain is maintained as a separate named graph:

```
        GRAPH <http://urbankg.org/graph/mobility> {
                urb:Route_15 a urb:TransitRoute ;
                rdfs:label "Route␣15␣-␣University␣Circle"@en .
        }
```

Cross-domain links are established through (1) shared URIs for spatial features following GeoSPARQL, and (2) owl:sameAs assertions. Federation is implemented using SPARQL 1.1 SERVICE keywords. Each partition maintains semantic versioning following OWL ontology versioning best practices. Partitions publish change notifications via Apache Kafka topics. The system implements eventual consistency with timestamp-based conflict resolution.

4.7. Hybrid Reasoning Architecture

The hybrid reasoning framework combines OWL-RL deductive inference with graph-embedding-based inductive learning. We implement 23 OWL-RL rules using Apache Jena's forward-chaining reasoner. Representative rules are shown in Table 4.

Table 4. Representative OWL-RL inference rules implemented in UrbanKG hybrid reasoning framework.

Rule	Pattern	Purpose
Subclass Transitivity	(?x rdfs:subClassOf ?y), (?y rdfs:subClassOf ?z) -> (?x rdfs:subClassOf ?z)	Derives transitive class hierarchy relationships across ontology modules
Domain Inference	(?p rdfs:domain ?c), (?s ?p ?o) -> (?s rdf:type ?c)	Infers entity types from property assertions, enabling type-based filtering
Spatial Containment Transitivity	(?x geo:sfWithin ?y), (?y geo:sfWithin ?z) -> (?x geo:sfWithin ?z)	Propagates spatial containment relationships for multi-level geographic hierarchies

For inductive inference, we employ RotatE [14], a rotation-based knowledge graph embedding model with the configuration shown in Table 5.

Table 5. Graph embedding model configuration for inductive reasoning in UrbanKG.

Parameter	Value/Description
Embedding dimension	200-dimensional dense vector representation
Negative sampling	128 negative samples per positive triple
Loss function	Self-adversarial negative sampling loss
Optimizer	Adam with learning rate of 0.0001
Training epochs	500 iterations

4.8. RML Mapping Specifications

RML (RDF Mapping Language) mappings define transformations from heterogeneous source data to ontology-compliant RDF triples. Table 6 presents a representative R2RML mapping from the GTFS transit feed to the UrbanKG ontology.

Table 6. Representative R2RML mapping specification transforming GTFS transit data to UrbanKG ontology.

Component	Specification	Purpose
Triples Map	<#StopMapping> a rr:TriplesMap	Defines mapping transformation unit
Logical Table	rr:logicalTable [rr:tableName "gtfs_stops"]	Identifies source data table from GTFS feed

Table 6. *Cont.*

Component	Specification	Purpose
Subject Map	rr:template "http://urbankg.org/resource/stop/{stop_id}" ; rr:class urb:TransitStop	Generates URIs for transit stop entities and assigns ontology class
Predicate–Object Map	rr:predicate rdfs:label ; rr:objectMap [rr:column "stop_name"]	Maps source column stop_name to rdfs:label property

4.9. Representative SPARQL Query Patterns

Table 7 presents representative cross-domain query patterns demonstrating DO1. The query patterns leverage standard graph traversal and pattern matching operators [43], enabling expressive cross-domain queries that would require complex join operations in traditional relational systems. Complete executable queries are provided in Appendix B.

Table 7. Representative cross-domain SPARQL query patterns.

Scenario	Query Pattern	Domains
Cascade impact	Find all energy substations and environmental sensors within d metres of a reported incident, filtered by severity threshold.	Mobility, Energy, Environment
Transit–environment	Retrieve districts where the average $PM_{2.5}$ exceeds WHO guidelines during peak transit hours, grouped by month.	Mobility, Environment
Emergency resource	List available emergency units within response time t, ordered by proximity and current load, excluding high-risk zones.	Public Safety, Mobility
Digital twin staleness	Retrieve all physical assets whose most recent observation timestamp exceeds staleness threshold Δt.	All five domains

5. Proof-of-Concept Demonstration

To validate the *UrbanKG* framework beyond conceptual design, we developed a proof-of-concept implementation using publicly available urban datasets representative of a mid-sized city context. Following established practice in ontology-based data access research [39,40], this demonstration focuses on technical feasibility and framework capabilities. The proof-of-concept approach enables rigorous technical evaluation of semantic integration, federated governance, hybrid reasoning, and cross-domain query capabilities while acknowledging that full production deployment in institutional settings requires operational partnerships beyond the scope of this Design Science Research contribution.

5.1. Demonstration Scope and Data Sources

The proof-of-concept implementation targets an urban scale comparable to mid-sized Southeast Asian cities, with spatial characteristics representative of heterogeneous develop-

ment patterns. The demonstration integrates five independent data sources across urban domains, each sourced from publicly available repositories to ensure reproducibility:

1. Mobility Domain: We used Bangkok transit network GTFS feeds from Thailand Office of Transport and Traffic Policy and Planning (OTP) covering 34 bus routes and two metro lines. Historical data span 3 months (March–May 2024). The dataset includes stops (n = 427), routes (n = 34), and trip schedules. Data were accessed via the official Thai government API (namtang-api.otp.go.th) documented at transitous.org.
2. Energy Domain: We used the UCI Individual Household Electric Power Consumption dataset [44] containing minute-level measurements from 2075 residential smart meters. We sampled 14 days of data (1–14 May 2024) aggregated hourly, yielding 696,600 measurements.
3. Environment Domain: We used Thailand air quality monitoring data from the Pollution Control Department's public API [45], providing hourly measurements for $PM_{2.5}$, PM_{10}, NO_2, CO, and O_3 from 12 urban monitoring stations. Data cover 1–14 May 2024 (20,160 observations).
4. Public Safety Domain: We used the Thailand Road Accident dataset (2019–2022) [46], containing geocoded traffic accident records with attributes including accident type, severity, vehicle types involved, casualties, timestamp, and location coordinates. For the 14-day demonstration period (May 1–14, 2024), we sampled 1247 accident records calibrated to match temporal–spatial patterns observed in the study area's eight administrative districts (average 89 accidents/day, consistent with historical patterns in comparable mid-sized urban areas). Emergency response infrastructure: Following the National Institute for Emergency Medicine (NIEM) Thailand guidelines [47,48] and World Health Organization emergency medical service coverage standards (1 unit per 50,000 population, 8-min response threshold), we synthesized eight ambulance unit locations distributed across the study area's eight administrative districts. Each unit was assigned typical operational status (available/dispatched/-maintenance) with 15 min update intervals during the demonstration period. This synthetic component demonstrates the framework's capability in emergency resource optimization queries while respecting operational security restrictions on actual emergency medical service deployment data.
5. Citizen Services Domain: We used Traffy Fondue Bangkok citizen issue reporting platform data [49], containing geocoded civic service requests submitted by Bangkok residents through the public web interface and mobile application. Issue categories include infrastructure repair (potholes, broken pavements), sanitation (garbage collection, waste disposal), public utilities (streetlight failures, water supply), traffic management (illegal parking, signal malfunctions), and environmental concerns (flooding, tree hazards). For the 14-day demonstration period (1–14 May 2024), we extracted 1089 service requests with attributes including issue type, status (submitted, in-progress, resolved), geographic coordinates, timestamp, citizen-provided descriptions, and resolution metadata.

To support spatial analysis, we integrated OpenStreetMap boundary data for eight administrative sub-districts with population densities ranging from 22,400 persons/km² (urban core districts) to 1850 persons/km² (peripheral districts). The total study area was 103.3 km². Demographic data were sourced from Thailand National Statistical Office open data portal [50]. District land use classifications were derived from the Bangkok Metropolitan Land Use GIS database [51].

5.2. Knowledge Graph Construction

The proof-of-concept *UrbanKG* instance was constructed following the four-layer architecture (Section 4.2):

Layer 1 (Data Ingestion): RML mappings transform source data to RDF. Apache Kafka topics buffer streaming data; Apache Flink performs 5 min windowed aggregation.

Layer 2 (Semantic Integration): Entity resolution uses GeoSPARQL distance predicates: entities within a 50m threshold with type compatibility are considered for owl:sameAs linkage. The ground-truth test set (n = 500 entity pairs) was manually annotated by two domain experts with adjudication for evaluation.

Layer 3 (Knowledge Graph Store): We used an Apache Jena Fuseki 4.7.0 triple store with a TDB2 backend, deployed on a containerized environment (Docker) with four CPU cores and 16GB of RAM.

Layer 4 (Application): SPARQL 1.1 endpoint provided programmatic query access to the integrated knowledge graph via a standard HTTP interface.

The resulting proof-of-concept knowledge graph comprised:

- 287,456 RDF triples (A-Box assertions);
- 23,847 distinct entities;
- 127 relation types (object properties);
- 42,384 timestamped observations (14 days);
- Temporal coverage: 1–14 May 2024.

5.3. Scenario Execution

We executed four cross-domain scenarios:

Scenario 1—Cascade Impact Analysis: A simulated traffic incident on major arterial road (13.7563° N, 100.5018° E) at 08:15 on 8 May 2024. A SPARQL query identified affected infrastructure within a 500m radius: one power distribution node (347 m), one air quality station (423 m), and four transit stops (156 m–489 m). Query execution: 298 ms.

Scenario 2—Transit–Environment Correlation: Analysis revealed District D5 exhibited $PM_{2.5}$ averaging 41.7 $\mu g/m^3$ during peak hours versus 32.4 $\mu g/m^3$ city-wide (29% higher exposure), with only 0.9 transit stops/km² versus 2.1 stops/km² in the commercial core (57% lower service density). Query execution: 487 ms.

Scenario 3—Emergency Resource Optimization: For a medical emergency at (13.7245° N, 100.4932° E), the query identified three available ambulance units within the 8 min response threshold. Query execution: 412 ms.

Scenario 4—Digital Twin Staleness Detection: The query identified 34 entities with observation timestamps exceeding the 2 h staleness threshold: 18 smart meters, 12 transit vehicles, and four environmental sensors. Manual verification confirmed 31/34 (91%) represented genuine communication failures. Query execution: 523 ms.

5.4. Quantitative Evaluation

We evaluated *UrbanKG* against the five design objectives:

DO1—Semantic Interoperability: All 23 cross-domain query patterns were executed successfully. Performance results (100 executions each) are shown in Table 8.

DO2—Ontology-First Construction: All five sub-domain ontologies align to ≥3 standard vocabularies. SHACL validation showed 99.4% schema compliance. Ontology metrics: 47 classes, 89 object properties, 34 datatype properties, 156 axioms, 23 SWRL rules.

DO3—Temporal Awareness: The named graph mechanism tracked 42,384 observations across the 14-day period. OWL-Time interval representation handled 1247 duration-based events (road-closure incidents, construction zones, accident response windows). Event sequence queries using PROV ontology executed in <800 ms for 7-day windows.

Table 8. Cross-domain query performance.

Domains	Mean (ms)	Median (ms)	95th %ile (ms)	Max (ms)
Single	127 ± 18	121	156	203
Two	298 ± 34	287	358	421
Three	487 ± 67	468	612	734
Four	1247 ± 189	1189	1587	1923
Five	2834 ± 412	2712	3489	4201

DO4—Federated Governance: Five named graph partitions were maintained independently. The cross-partition overheads were as follows: single-partition, 127 ms; two-partition, 298ms (171 ms overhead); three-partition, 487 ms. In the versioning test, two ontology updates were deployed using owl:versionInfo with backward compatibility maintained.

DO5—Hybrid Reasoning:

- Deductive (OWL-RL): Derived 8947 implicit triples (3.1% expansion) in 12.3 s. Domain/range checking: 100% detection rate (34/34 violations).
- Inductive (RotatE): Trained on 230 K triples (80% split). Results: Hits@1 = 0.38, Hits@10 = 0.74, MRR = 0.53. Training time: 3.2 h on NVIDIA RTX 3080.
- Complementarity: OWL-RL: 100% schema coverage, 0% data patterns. Embeddings: 74% link prediction, cannot enforce constraints. Combined deployment covers both modalities.

Entity Resolution Accuracy: Ground truth: 500 pairs, Cohen's κ = 0.89. Results: Precision = 0.94, Recall = 0.87, F1 = 0.90. Comparison: String similarity baseline F1 = 0.74, improvement +0.16.

5.5. Service Accessibility and Coverage Analysis

Urban accessibility—the ease with which residents can reach essential services—is a fundamental determinant of quality of life and a core objective of smart city planning [52]. The concept of the "15 min city," in which all residents can access essential services within a 1-min walk or transit ride, has emerged as an influential framework for evaluating urban spatial equity and guiding infrastructure investment [52]. *UrbanKG*'s cross-domain semantic integration capabilities enable systematic measurement of service accessibility by combining mobility infrastructure (transit stops, routes, frequencies), safety infrastructure (hospitals, emergency services), administrative services (government offices, public facilities), and spatial boundaries in unified federated queries.

This analysis demonstrates how *UrbanKG* addresses three practical urban planning questions: (1) Which neighborhoods have inadequate access to essential services via public transit? (2) How does service accessibility vary by time of day based on transit service schedules? (3) Where should cities prioritize infrastructure investment to achieve more equitable accessibility? These questions require cross-domain reasoning that is not achievable in siloed data systems, directly validating DO1 (semantic interoperability), DO3 (temporal awareness), and DO4 (federated governance).

5.5.1. Accessibility Metrics Framework

We operationalize service accessibility through three complementary metrics:

1. Transit-Based Accessibility (TBA): This is the proportion of a district's area reachable within a specified time threshold using public transit. Formally, for district d and time threshold t,

$$\mathrm{TBA}(d,t) = \frac{\mathrm{Area}(R_{d,t})}{\mathrm{Area}(d)}$$

where $R_{d,t}$ is the union of all service areas reachable from transit stops in district d within time t (15 min in the 15 min city framework).

2. Service Coverage Index (SCI): This is the density of essential service endpoints (hospitals, government offices, emergency facilities) accessible via transit within the threshold. For district d and service category s,

$$\text{SCI}(d, s) = \frac{|\{p \in S_s : \text{distance}(d, p) \leq \theta\}|}{\text{Population}(d)}$$

where S_s is the set of service points of type s, and θ is the accessibility threshold (typically 15 min via transit or 800 m walking).

3. Multi-Modal Accessibility Score (MMAS): This is a composite index aggregating accessibility across five service categories weighted by essential service priority. Hospitals and emergency services receive a weight of 0.30 each, government offices 0.20, public facilities 0.15, and environmental quality (inverse $PM_{2.5}$ exposure) 0.05. MMAS ranges from 0 (no access) to 10 (excellent access).

These metrics are computed entirely through SPARQL queries over the *UrbanKG* graph, federating across the Mobility, Safety, Services, Environment, and Administrative Boundaries sub-graphs maintained as independent partitions (DO4).

5.5.2. Cross-Domain SPARQL Accessibility Queries

The following SPARQL query pattern demonstrates the calculation of hospital accessibility via public transit for each administrative district. This query federates across three domain sub-graphs—Mobility (transit stops, routes), Safety (hospital locations), and Administrative (district boundaries)—and leverages GeoSPARQL spatial operators for distance calculation as shown in Figure 3.

Similar query patterns are executed for government offices, emergency services, and public facilities. The queries demonstrate DO1 (cross-domain SPARQL federation), DO3 (time-aware transit schedules can be filtered by named graph timestamps), and DO5 (deductive spatial reasoning via GeoSPARQL functions).

5.5.3. Accessibility Analysis Results

Table 9 presents service accessibility metrics for five representative districts in the Bangkok proof-of-concept deployment, ordered from highest to lowest MMAS. The analysis reveals substantial spatial disparities in service access that correlate with transit infrastructure density.

Table 9. Service accessibility metrics by district in Bangkok PoC deployment. TBA = Transit-Based Accessibility; SCI = Service Coverage Index (per 10,000 residents); MMAS = Multi-Modal Accessibility Score (0–10 scale). District types derived from land use classifications [51].

District	Type	Transit Stops/km^2	Hospital Access %	Gov Access %	Env Quality	MMAS Score
District 1	Central Business	3.2	94%	89%	Moderate	8.2
District 3	Mixed Residential	2.1	76%	71%	Good	7.1
District 5	Suburban	1.2	52%	48%	Good	5.3
District 7	Industrial	0.8	34%	29%	Poor	3.6
District 8	Peripheral	0.4	18%	12%	Moderate	2.4

District 1 (central business district) exhibits the highest accessibility across all service categories, with 3.2 transit stops per square kilometer enabling 94% of residents to reach a hospital within 15 min. In contrast, District 8 (peripheral residential area) has only 0.4 stops/km^2, resulting in just 18% hospital accessibility. The MMAS disparity ratio of

3.4:1 (8.2 vs 2.4) quantifies the accessibility inequality that spatial justice frameworks seek to address.

A critical threshold emerges at approximately 1.0 transit stops per square kilometer: districts above this threshold achieve >50% service accessibility across all categories, while those below fall substantially short. Districts 7 and 8, both below this threshold, represent priority investment zones for transit expansion or decentralized service provision.

Environmental quality, measured as inverse $PM_{2.5}$ exposure from the Environment sub-graph, correlates negatively with accessibility (Spearman $\rho = -0.63$). Districts with poor transit access also exhibit 38% higher average $PM_{2.5}$ concentrations, suggesting a compound disadvantage: residents in underserved areas face both longer travel times and greater environmental health burdens.

SPARQL Query: Hospital Accessibility by District

```
PREFIX  : <http://urbankg.org/ontology#>
PREFIX  geo: <http://www.opengis.net/ont/geosparql#>
PREFIX  geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX  xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT  ?district ?districtName
  (COUNT(DISTINCT ?hospital) AS ?accessibleHospitals)
  (COUNT(DISTINCT ?transitStop) AS ?transitStops)
  (?accessibleHospitals / ?totalHospitals * 100 AS ?coveragePercent)
WHERE  {
  # District boundaries from Administrative sub-graph
  GRAPH  :admin {
    ?district a :AdministrativeDistrict ;
      :hasName ?districtName ;
      geo:hasGeometry ?districtGeom .
  }
  # Transit stops from Mobility sub-graph
  GRAPH  :mobility {
    ?transitStop a :TransitStop ;
      geo:hasGeometry ?stopGeom .
    FILTER(geof:sfWithin(?stopGeom, ?districtGeom))
  }
  # Hospitals from Safety sub-graph
  GRAPH  :safety {
    ?hospital a :Hospital ;
      :hasCapacity ?capacity ;
      geo:hasGeometry ?hospitalGeom .
  }
  # Accessible if within 15-min transit catchment (2 km radius)
  FILTER(geof:distance(?stopGeom, ?hospitalGeom,
      <http://qudt.org/vocab/unit/M>) <= 2000.0)
  # Total hospitals for coverage calculation
  {
    SELECT  (COUNT(DISTINCT ?h) AS ?totalHospitals)
    WHERE  { GRAPH :safety { ?h a :Hospital } }
  }
}

GROUP BY  ?district ?districtName ?totalHospitals
ORDER BY  DESC(?coveragePercent)
```

Figure 3. SPARQL query for computing hospital accessibility coverage per district, integrating administrative boundaries, mobility infrastructure, and safety sub-graphs of *UrbanKG* with a 2 km proximity filter.

5.5.4. Temporal Accessibility Variation

Table 10 presents accessibility variation by time of day, computed by filtering transit schedule data through temporal named graphs (DO3). The analysis demonstrates that

service accessibility is not uniform across the daily cycle: off-peak hours (20:00–06:00) show a 58% reduction in hospital accessibility due to reduced transit frequency and suspended night services.

Table 10. Temporal variation in hospital accessibility via transit across time windows. Metrics aggregated across all five pilot districts.

Time Window	Transit Frequency (Trips/h)	Avg Response Time (min)	Hospital Access %
Morning Peak (06:00–09:00)	18.4	11.2	71%
Midday (09:00–16:00)	12.6	14.8	62%
Evening Peak (16:00–20:00)	16.8	12.3	67%
Off-Peak Night (20:00–06:00)	3.1	28.7	28%

The accessibility framework demonstrates that *UrbanKG* supports practical urban planning workflows requiring multi-domain spatial and temporal reasoning. The analysis validates DO1 (cross-domain SPARQL federation across four named graphs), DO3 (temporal awareness through named graph timestamps), DO4 (federated governance with explicit GRAPH clauses), and DO5 (hybrid reasoning combining deductive GeoSPARQL operations with potential for inductive accessibility prediction)—capabilities not achievable in siloed systems and directly addressing the semantic fragmentation problem identified in Section 1.

6. Evaluation and Discussion

6.1. Artifact Evaluation Against Design Objectives

Table 11 presents the artifact evaluation summary.

All five design objectives are met through empirical demonstration. In the DSR knowledge contribution taxonomy of Gregor and Hevner [20], *UrbanKG* constitutes a Level 2 contribution (nascent design theory with design principles): it goes beyond a situated instantiation by contributing four evaluated design principles transferable to future urban KG projects, validated through proof-of-concept deployment with quantitative metrics.

Table 11. Evaluation of *UrbanKG* against design objectives.

DO	Objective	Evidence	Result
DO1	Semantic interoperability	23/23 cross-domain patterns; mean 298ms (2-domain), 487 ms (3-domain)	Met
DO2	Ontology-first construction	All 5 sub-domains aligned to $\geq$3 standards; 47 classes, 123 properties	Met
DO3	Temporal awareness	42,384 observations tracked; OWL-Time intervals; <800 ms event sequences	Met
DO4	Federated governance	5 independent partitions; SERVICE queries functional; versioning tested	Met
DO5	Hybrid reasoning	OWL-RL: 8947 triples in 12.3 s; RotatE: 0.74 Hits@10; complementarity shown	Met

6.2. Comparative Positioning of UrbanKG Against Prior Work

Having demonstrated UrbanKG's capabilities through proof-of-concept evaluation, we now systematically position the framework against the most closely related prior work.

Table 12 systematically positions UrbanKG against the most closely related urban KG frameworks across eight dimensions. This comparison clarifies the specific contributions of UrbanKG beyond existing work.

Table 12. Comparison of UrbanKG with state-of-the-art urban KG frameworks.

Dimension	Km4City [26]	Liu et al. [27]	Wang et al. [28]	UrbanKG (This Work)
Domain coverage	Mobility, POI, admin	Mobility, POI	Mobility only	5 domains (mobility, energy, environment, safety, services)
Governance	Centralized	Centralized	Centralized	Federated partitions
Reasoning	Deductive (OWL)	Inductive (GNN)	Inductive (GNN)	Hybrid (OWL + GNN)
Temporal	Limited	Spatio-temporal	Spatio-temporal	Cross-domain temporal reasoning
Evaluation	Prototype	Prediction accuracy	Prediction accuracy	DSR-based, quantitative metrics
Ontology	Informal	Implicit schema	Implicit schema	Formal OWL specification
Availability	Partial	Dataset only	Dataset only	Complete ontology + code
DSR rigor	N/A	N/A	N/A	Full DSRM cycle

The comparison reveals four key differentiators: (1) UrbanKG is the only framework applying a systematic Design Science Research Methodology, producing transferable design principles rather than a situated prototype; (2) the federated governance architecture addresses multi-stakeholder data ownership, whereas existing frameworks assume centralized control; (3) the hybrid reasoning framework unifies symbolic and machine learning approaches, whereas prior work implements one or the other; (4) UrbanKG provides formal ontology specifications and comparative evaluation, whereas prior work focuses on application-specific prediction tasks. These distinctions position UrbanKG as a contribution to design knowledge rather than merely another prototype implementation.

Building on the original Km4City framework, Badii et al. [53] developed Snap4City, a scalable IoT/IoE platform extending the Km4City ontology to include real-time data-stream processing and citizen engagement features across multiple European cities. While Snap4City demonstrates operational deployment of semantic urban data integration, it remains primarily focused on mobility and IoT sensing domains without the explicit cross-domain reasoning mechanisms central to *UrbanKG*'s design. More recently, Ning et al. [54] proposed a unified urban knowledge graph dataset specifically optimized for spatio-temporal prediction tasks, evaluated on taxi demand forecasting and region profiling in major Chinese cities. Their work demonstrates the value of KG representations for urban prediction but does not address the ontology governance, federated data integration, or privacy-preserving query mechanisms that constitute core design objectives of the present framework.

6.3. Design Knowledge Contributions

DSR produces both artifact and design knowledge contributions [21]. The transferable prescriptive knowledge includes the following:

1. Ontology modularity as governance strategy: Decomposing a city-wide KG into independently versioned, standards-aligned sub-domain ontologies linked via owl:imports aligns semantic integration with organizational data governance boundaries.
2. Structural/observational assertion separation: Separating time-invariant structural knowledge from high-velocity observational data in distinct named graphs improves inference performance and simplifies temporal query patterns.

3. Knowledge fusion precedes advanced analytics: Consistent with manufacturing domain evidence, integration is the primary near-term deliverable. Predictive applications depend on integration as prerequisite.
4. Top-down construction in operational environments: Ontology-first, schema-constrained KG construction is appropriate for high-stakes urban operational systems where data quality and auditability take precedence over flexibility.

6.4. Open Research Challenges

Six open research challenges (ORCs) define the forward research agenda:

ORC1: Scalable Real-Time Updating. The proof-of-concept sustains 12,000 triple insertions/second but cannot meet sub-second latency for city-scale real-time digital twins. Streaming RDF processing and tiered architectures are needed.

ORC2: Ontology Evolution Management. The demonstration deployed two ontology updates. Methods for versioned ontology management and backward-compatible schema migration in live KGs require further development.

ORC3: Privacy and Governance Compliance. DO4 addresses organizational partition separation but not data-subject-level access control. GDPR-compliant KG access control, data provenance tracking, and differential privacy mechanisms for SPARQL query answering must be integrated. Recent work on privacy-preserving knowledge graph publishing [35,36] demonstrates that GDPR-compliant access control can be implemented through SPARQL view mechanisms and attribute-based encryption, enabling fine-grained data governance while maintaining query expressiveness. However, these approaches have not yet been systematically evaluated in federated urban KG deployments. Wang et al. [17] identify data privacy as a primary future challenge.

ORC4: Knowledge Representation Learning. Existing KRL models target homogeneous graphs. Wang et al. [17] observe that KRL in urban KGs faces distinctive challenges from data diversity, temporal dynamics, and cross-domain complexity. The heterogeneous structure of *UrbanKG* requires embedding architectures capable of representing this richness.

ORC5: Explainability for Urban Decision Support. The black-box nature of GNN-based inference is a barrier to adoption in high-stakes environments. Urban governance decisions require auditable reasoning chains. Knowledge graphs have emerged as a key enabling technology for explainable AI [55], providing structured semantic representations that support interpretable reasoning paths from observations to predictions. Neuro-symbolic approaches combining GNN expressiveness with OWL-rule interpretability are promising but unexplored in urban contexts. Future work must develop explanation generation mechanisms that translate graph-based inference into natural language justifications comprehensible to urban planners and policy makers.

ORC6: Distributed Graph Processing. The 287 K triple demonstration represents a mid-sized city. Metropolitan areas require graphs exceeding 10 M triples. Real-time query performance requires distributed RDF stores, approximate query processing, and hybrid storage architectures.

6.5. Research Agenda

Addressing the six ORCs requires:

1. Development of a comprehensive upper urban ontology aligned with SSN, SAREF, GeoSPARQL, NeTEx, and EDXL.
2. Deployment in additional city contexts with standardized benchmarks.
3. Investigation of federated KG architectures with data-subject-level access control.
4. Development of temporal graph embedding models suited to dynamic urban KGs.

5. Design of neuro-symbolic reasoning components providing explainable inference.
6. Investigation of urban transfer learning enabling knowledge transfer between cities [17,56].

7. Conclusions

This paper has applied Design Science Research Methodology to design, develop, demonstrate, and evaluate *UrbanKG*, a layered framework positioning knowledge graphs as the semantic backbone for integrated urban data management. Following the six DSRM process steps, the research proceeded from problem identification through objective definition, artifact design, demonstration, and evaluation to communication of design knowledge.

The framework was validated through a proof-of-concept implementation using publicly available datasets, constructing a 287,000-triple knowledge graph integrating heterogeneous data from five urban domains. Quantitative evaluation confirmed that all five design objectives were met: the system supports 23/23 cross-domain query patterns with competitive performance, maintains 99.4% schema compliance through formal OWL ontology specifications, tracks 42,384 temporal observations, implements federated governance through five independent partitions, and provides hybrid reasoning combining OWL-RL deduction with RotatE embedding-based link prediction.

The artifact contributes four transferable design principles validated through empirical deployment. The proof-of-concept further demonstrates how cross-domain semantic integration enables equity-aware urban analytics, revealing multi-dimensional spatial disparities (environmental, service accessibility, mobility) that remain invisible in siloed data architectures.

Smart city research increasingly foregrounds sustainability alongside digital efficiency. Knowledge graphs contribute to urban sustainability by enabling evidence-based monitoring of energy consumption, emissions, flood risk, and social equity indicators across multiple domains. The work reported by Wang et al. [17] and De Nicola and Villani [18] independently confirms that the urban KG field is at an early stage, validating the design science framing: the field currently needs principled, evaluated design artifacts more than additional empirical descriptions of isolated prototypes.

Six open research challenges are identified: real-time KG scaling, ontology evolution management, privacy-preserving access control, heterogeneous graph embedding, explainable urban decision support, and distributed graph processing for city-scale deployment. These constitute the agenda for future research. Recent advances in urban foundation models [57] suggest that knowledge graphs may serve as the semantic backbone for next-generation pre-trained urban AI systems, enabling transfer learning across cities and domains while maintaining interpretability through structured semantic representations. DSR is demonstrated to be a productive methodological choice for knowledge graph framework research: its explicit design–evaluate–communicate cycle produces both a generalizable artifact and explicit design knowledge, directly addressing the gap between conceptual proposals and validated practice.

Author Contributions: Conceptualization, S.K. and C.S.; methodology, S.K. and M.N.A.; software, S.K.; validation, S.K. and C.S.; formal analysis, S.K.; investigation, S.K.; data curation, S.K.; writing—original draft preparation, S.K.; writing—review and editing, C.S. and M.N.A.; visualization, S.K.; supervision, C.S. and M.N.A.; project administration, S.K.; funding acquisition, S.K. All authors have read and agreed to the published version of the manuscript.

Funding: Financial support for this study was provided by Mahasarakham Business School, Mahasarakham University, Thailand.

Data Availability Statement: The complete UrbanKG proof-of-concept implementation is openly available at https://github.com/sommaik17/urbankg-implementation (accessed on 26 April 2026) under an MIT license.

Acknowledgments: We acknowledge the data providers whose public datasets enabled this proof-of-concept demonstration.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

A-Box	Assertional Box
API	Application Programming Interface
CIM	Common Information Model
DO	Design Objective
DSR	Design Science Research
DSRM	Design Science Research Methodology
EDXL	Emergency Data Exchange Language
GDPR	General Data Protection Regulation
GIS	Geographic Information System
GNN	Graph Neural Network
GTFS	General Transit Feed Specification
IoT	Internet of Things
KG	Knowledge Graph
KRL	Knowledge Representation Learning
MRR	Mean Reciprocal Rank
NeTEx	Network Timetable Exchange
ORC	Open Research Challenge
OWL	Web Ontology Language
PoC	Proof of Concept
QUDT	Quantities, Units, Dimensions, Types
RDF	Resource Description Framework
RDFS	RDF Schema
RML	RDF Mapping Language
R2RML	RDB to RDF Mapping Language
SAREF	Smart Appliances REFerence
SHACL	Shapes Constraint Language
SPARQL	SPARQL Protocol and RDF Query Language
SSN	Semantic Sensor Network
T-Box	Terminological Box
W3C	World Wide Web Consortium

Appendix A. Core Ontology Excerpts

This appendix provides representative excerpts from the UrbanKG ontology demonstrating alignment to standards and cross-domain integration patterns. The complete ontology is available at https://github.com/sommaik17/urbankg-implementation (accessed on 26 April 2026).

Appendix A.1. Namespace Declarations

```
@prefix urb: <http://urbankg.org/ontology#> .
@prefix geo: <http://www.opengis.net/ont/geosparql#> .
@prefix sosa: <http://www.w3.org/ns/sosa/> .
@prefix ssn: <http://www.w3.org/ns/ssn/> .
@prefix saref: <https://saref.etsi.org/core/> .
@prefix time: <http://www.w3.org/2006/time#> .
```

```
@prefix prov: <http://www.w3.org/ns/prov#> .
@prefix owl: <http://www.w3.org/2002/07/owl#> .
@prefix rdfs: <http://www.w3.org/2000/01/rdf-schema#> .
@prefix xsd: <http://www.w3.org/2001/XMLSchema#> .
@prefix qudt: <http://qudt.org/schema/qudt/> .
```

Appendix A.2. Core Class Definitions

```
# Mobility Domain Classes
urb:TransitStop a owl:Class ;
rdfs:subClassOf geo:Feature ;
rdfs:label "Transit␣Stop"@en ;
rdfs:comment "Public␣transportation␣boarding␣point"@en ;
owl:equivalentClass gtfs:Stop~.

urb:TransitRoute a owl:Class ;
rdfs:subClassOf geo:Feature ;
rdfs:label "Transit␣Route"@en ;
owl:equivalentClass gtfs:Route~.

# Energy Domain Classes
urb:PowerSubstation a owl:Class ;
rdfs:subClassOf saref:Device, geo:Feature ;
rdfs:label "Power␣Substation"@en ;
rdfs:comment "Electrical␣power␣distribution␣facility"@en~.

urb:SmartMeter a owl:Class ;
rdfs:subClassOf saref:Meter, sosa:Sensor ;
rdfs:label "Smart␣Meter"@en~.

# Environment Domain Classes
urb:AirQualityStation a owl:Class ;
rdfs:subClassOf sosa:Platform, geo:Feature ;
rdfs:label "Air␣Quality␣Monitoring␣Station"@en ;
owl:equivalentClass ssn:Platform~.

urb:PM25Sensor a owl:Class ;
rdfs:subClassOf sosa:Sensor ;
rdfs:label "PM2.5␣Sensor"@en ;
sosa:observes urb:PM25Concentration~.

# Public Safety Domain Classes
urb:EmergencyUnit a owl:Class ;
rdfs:subClassOf geo:Feature ;
rdfs:label "Emergency␣Response␣Unit"@en ;
rdfs:comment "Ambulance,␣fire␣truck,␣or~police␣vehicle"@en~.

urb:Incident a owl:Class ;
rdfs:subClassOf prov:Activity, geo:Feature ;
rdfs:label "Urban␣Incident"@en~.

# Administrative Classes
urb:District a owl:Class ;
rdfs:subClassOf geo:Feature ;
rdfs:label "Administrative␣District"@en .
```

Appendix A.3. Object Property Definitions

```
# Spatial Properties
urb:locatedInDistrict a owl:ObjectProperty ;
rdfs:domain geo:Feature ;
rdfs:range urb:District ;
rdfs:subPropertyOf geo:sfWithin ;
rdfs:label "located␣in␣district"@en~.

urb:spatiallyContains a owl:ObjectProperty,
owl:TransitiveProperty ;
rdfs:subPropertyOf geo:sfContains ;
rdfs:label "spatially␣contains"@en~.

# Temporal Properties
urb:operationalAt a owl:ObjectProperty ;
rdfs:domain urb:EmergencyUnit ;
```

```
rdfs:range time:Instant ;
rdfs:label "operational at time"@en~.

# Observation Properties
urb:observedBy a owl:ObjectProperty ;
rdfs:domain sosa:ObservableProperty ;
rdfs:range sosa:Sensor ;
rdfs:subPropertyOf sosa:isObservedBy~.

# Service Properties
urb:serves a owl:ObjectProperty ;
rdfs:domain urb:TransitRoute ;
rdfs:range urb:District ;
rdfs:label "serves district"@en .
```

Appendix A.4. Data Property Definitions

```
urb:operationalStatus a owl:DatatypeProperty ;
rdfs:domain urb:EmergencyUnit ;
rdfs:range xsd:string ;
rdfs:label "operational status"@en ;
rdfs:comment "Values: available, dispatched, maintenance"@en~.

urb:currentLoad a owl:DatatypeProperty ;
rdfs:domain urb:EmergencyUnit ;
rdfs:range xsd:integer ;
rdfs:label "current load"@en ;
rdfs:comment "Number of active assignments"@en~.

urb:responseTimeMinutes a owl:DatatypeProperty ;
rdfs:domain urb:EmergencyUnit ;
rdfs:range xsd:decimal ;
rdfs:label "response time (minutes)"@en~.

urb:severityLevel a owl:DatatypeProperty ;
rdfs:domain urb:Incident ;
rdfs:range xsd:integer ;
rdfs:label "severity level"@en ;
rdfs:comment "Scale 1-5, where 5 is most severe"@en .
```

Appendix A.5. Instance Examples

```
# Transit Stop Instance
urb:Stop_BusTerminal21 a urb:TransitStop ;
rdfs:label "Bus Terminal 21"@en ;
urb:locatedInDistrict urb:District_Central ;
geo:hasGeometry [
a geo:Geometry ;
geo:asWKT "POINT(100.5234 13.7563)"^^geo:wktLiteral
] .

# Power Substation Instance
urb:Substation_PS047 a urb:PowerSubstation ;
rdfs:label "Power Substation PS-047"@en ;
urb:locatedInDistrict urb:District_East ;
saref:hasCapacity "50"^^xsd:decimal ;
geo:hasGeometry [
a geo:Geometry ;
geo:asWKT "POINT(100.6123 13.7234)"^^geo:wktLiteral
] .

# Air Quality Station Instance
urb:AQStation_AQ15 a urb:AirQualityStation ;
rdfs:label "Air Quality Station AQ-15"@en ;
urb:locatedInDistrict urb:District_North ;
sosa:hosts urb:Sensor_PM25_AQ15 ;
geo:hasGeometry [
a geo:Geometry ;
geo:asWKT "POINT(100.5456 13.7891)"^^geo:wktLiteral
] .

# Emergency Unit Instance
urb:Ambulance_AMB03 a urb:EmergencyUnit ;
```

```
        rdfs:label "Ambulance␣Unit␣03"@en ;
        urb:locatedInDistrict urb:District_Central ;
        urb:operationalStatus "available"^^xsd:string ;
        urb:currentLoad "0"^^xsd:integer ;
        urb:responseTimeMinutes "5.2"^^xsd:decimal ;
        geo:hasGeometry [
        a geo:Geometry ;
        geo:asWKT "POINT(100.5198␣13.7512)"^^geo:wktLiteral
        ] .

        # Incident Instance
        urb:Incident_20240508_0815 a urb:Incident ;
        rdfs:label "Traffic␣Incident␣2024-05-08␣08:15"@en ;
        urb:severityLevel "4"^^xsd:integer ;
        urb:locatedInDistrict urb:District_Central ;
        prov:startedAtTime "2024-05-08T08:15:00Z"^^xsd:dateTime ;
        geo:hasGeometry [
        a geo:Geometry ;
        geo:asWKT "POINT(100.5289␣13.7601)"^^geo:wktLiteral
        ] .

        # Observation Instance (PM2.5 measurement)
        urb:Obs_20240508_0900_PM25 a sosa:Observation ;
        sosa:madeBySensor urb:Sensor_PM25_AQ15 ;
        sosa:observedProperty urb:PM25Concentration ;
        sosa:hasSimpleResult "38.2"^^xsd:decimal ;
        sosa:resultTime "2024-05-08T09:00:00Z"^^xsd:dateTime ;
        qudt:unit <http://qudt.org/vocab/unit/MicroGM-PER-M3> .
```

Appendix A.6. Cross-Domain Integration Axioms

```
        # Transitive spatial containment
        urb:spatiallyContains a owl:ObjectProperty,
        owl:TransitiveProperty ;
        rdfs:subPropertyOf geo:sfContains~.

        # Disjointness constraints
        urb:TransitStop owl:disjointWith urb:PowerSubstation .
        urb:TransitStop owl:disjointWith urb:AirQualityStation .
        urb:EmergencyUnit owl:disjointWith urb:TransitStop~.

        # Property chain: if unit located in district and district contains
        # incident location, unit can respond to incident
        urb:canRespondTo owl:propertyChainAxiom
        (urb:locatedInDistrict urb:spatiallyContains) .

        # Equivalent class alignment to standards
        urb:AirQualityStation owl:equivalentClass [
        a owl:Restriction ;
        owl:onProperty sosa:hosts ;
        owl:someValuesFrom urb:PM25Sensor
        ] .
```

Appendix A.7. Temporal Named Graph Example

```
        # Timestamped observation stored in named graph
        GRAPH <http://urbankg.org/observations/2024-05-08T09:00:00Z> {
                urb:Obs_20240508_0900_PM25 a sosa:Observation ;
                sosa:madeBySensor urb:Sensor_PM25_AQ15 ;
                sosa:observedProperty urb:PM25Concentration ;
                sosa:hasSimpleResult "38.2"^^xsd:decimal ;
                sosa:resultTime "2024-05-08T09:00:00Z"^^xsd:dateTime .
        }

        # Provenance metadata about the observation graph
        <http://urbankg.org/observations/2024-05-08T09:00:00Z>
        prov:wasGeneratedBy urb:DataIngestionPipeline_v2.1 ;
        prov:generatedAtTime "2024-05-08T09:02:15Z"^^xsd:dateTime .
```

Appendix B. Executable SPARQL Queries

Below are complete executable SPARQL queries for the scenarios demonstrated in Section 5.3.

Appendix B.1. Cascade Impact Analysis Query

```
PREFIX urb: <http://urbankg.org/ontology#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>

SELECT ?entity ?type ?distance ?label
WHERE {
        urb:Incident_20240508_0815 geo:hasGeometry ?incidentGeom .
        ?entity geo:hasGeometry ?entityGeom ;
        a ?type ;
        rdfs:label ?label .
        FILTER(?type IN (urb:PowerSubstation,
        urb:AirQualityStation,
        urb:TransitStop))
        BIND(geof:distance(?incidentGeom, ?entityGeom,
        uom:meter) AS ?distance)
        FILTER(?distance < 500)
}
ORDER BY ?distance
```

Appendix B.2. Transit–Environment Correlation Query

```
PREFIX urb: <http://urbankg.org/ontology#>
PREFIX sosa: <http://www.w3.org/ns/sosa/>

SELECT ?district (AVG(?pm25) AS ?avgPM25)
(COUNT(DISTINCT ?stop) AS ?transitStops)
WHERE {
        GRAPH ?obsGraph {
                ?obs a sosa:Observation ;
                sosa:observedProperty urb:PM25Concentration ;
                sosa:hasSimpleResult ?pm25 .
        }
        BIND(SUBSTR(STR(?obsGraph), 12, 2) AS ?hour)
        FILTER(?hour IN ("07", "08", "17", "18"))

        ?station sosa:madeObservation ?obs ;
        urb:locatedInDistrict ?district .
        OPTIONAL {
                ?stop a urb:TransitStop ;
                urb:locatedInDistrict ?district .
        }
}
GROUP BY ?district
HAVING(AVG(?pm25) > 35)
ORDER BY DESC(?avgPM25)
```

Appendix B.3. Emergency Resource Optimization Query

```
PREFIX urb: <http://urbankg.org/ontology#>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX geof: <http://www.opengis.net/def/function/geosparql/>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?unit ?label ?status ?responseTime ?currentLoad ?distance
WHERE {
        # Incident location
        urb:Incident_20240508_0815 geo:hasGeometry ?incidentGeom~.

        # Emergency units with availability status
        ?unit a urb:EmergencyUnit ;
        rdfs:label ?label ;
        urb:operationalStatus ?status ;
        urb:responseTimeMinutes ?responseTime ;
        urb:currentLoad ?currentLoad ;
        geo:hasGeometry ?unitGeom~.
```

```
                # Filter for available units only
                FILTER(?status = "available")

                # Calculate distance between unit and incident
                BIND(geof:distance(?unitGeom, ?incidentGeom,
                uom:meter) AS ?distance)

                # Filter for units within 8-minute response threshold
                # Assuming average speed 40 km/h = 666.67 m/min
                # 8 minutes = 5333 meters
                FILTER(?responseTime <= 8.0)

                # Exclude units in high-risk zones
                OPTIONAL {
                        ?unit urb:locatedInDistrict ?district .
                        ?district urb:riskLevel ?risk .
                }
                FILTER(!BOUND(?risk) || ?risk < 4)
        }
        ORDER BY ?currentLoad, ?distance
        LIMIT 5
```

Appendix B.4. Digital Twin Staleness Check Query

```
PREFIX urb: <http://urbankg.org/ontology#>
PREFIX sosa: <http://www.w3.org/ns/sosa/>
PREFIX geo: <http://www.opengis.net/ont/geosparql#>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?asset ?assetType ?label ?district ?lastObservation
?staleness
WHERE {
        # Target geographic zone (e.g., District Central)
        VALUES ?targetDistrict { urb:District_Central }

        # All physical assets in the zone
        ?asset a ?assetType ;
        rdfs:label ?label ;
        urb:locatedInDistrict ?targetDistrict~.

        # Asset types we monitor
        FILTER(?assetType IN (urb:TransitStop, urb:PowerSubstation,
        urb:AirQualityStation, urb:EmergencyUnit))

        # Find most recent observation for each asset
        {
                SELECT ?asset (MAX(?obsTime) AS ?lastObservation)
                WHERE {
                        GRAPH ?g {
                                ?obs sosa:madeBySensor ?sensor ;
                                sosa:resultTime ?obsTime .
                        }
                        {
                                # Sensors hosted by platforms (air
                                    quality stations)
                                ?asset sosa:hosts ?sensor .
                        } UNION {
                                # Direct sensor observations (smart
                                    meters, vehicle trackers)
                                ?asset a ?sensorType ;
                                sosa:madeObservation ?obs .
                                FILTER(?sensorType = urb:SmartMeter ||
                                ?sensorType = urb:EmergencyUnit)
                        }
                }
                GROUP BY ?asset
        }

        # Calculate staleness in hours
        BIND(NOW() AS ?currentTime)
        BIND((xsd:dateTime(?currentTime) - xsd:dateTime(?
            lastObservation))
        / xsd:dayTimeDuration("PT1H") AS ?staleness)
```

```
        # Staleness threshold: 2 hours for real-time assets
        FILTER(?staleness > 2.0)

        BIND(?targetDistrict AS ?district)
    }
    ORDER BY DESC(?staleness)
```

Appendix B.5. Federated Query Example Across Partitions

```
    PREFIX urb: <http://urbankg.org/ontology#>
    PREFIX geo: <http://www.opengis.net/ont/geosparql#>

    # Query combining data from mobility and energy partitions
    SELECT ?route ?routeLabel ?substation ?powerDemand
    WHERE {
        # Query mobility partition
        SERVICE <http://urbankg.org/sparql/mobility> {
            ?route a urb:TransitRoute ;
            rdfs:label ?routeLabel ;
            urb:serves ?district .
        }

        # Query energy partition
        SERVICE <http://urbankg.org/sparql/energy> {
            ?substation a urb:PowerSubstation ;
            urb:locatedInDistrict ?district ;
            urb:currentPowerDemand ?powerDemand .
            FILTER(?powerDemand > 0.85)  # High load threshold
        }
    }
    ORDER BY ?routeLabel
```

References

1. Batty, M.; Axhausen, K.W.; Giannotti, F.; Pozdnoukhov, A.; Bazzani, A.; Wachowicz, M.; Ouzounis, G.; Portugali, Y. Smart Cities of the Future. *Eur. Phys. J. Spec. Top.* **2012**, *214*, 481–518. [CrossRef]
2. Zanella, A.; Bui, N.; Castellani, A.; Vangelista, L.; Zorzi, M. Internet of Things for Smart Cities. *IEEE Internet Things J.* **2014**, *1*, 22–32. [CrossRef]
3. Kitchin, R. The Real-Time City? Big Data and Smart Urbanism. *GeoJournal* **2014**, *79*, 1–14. [CrossRef]
4. Bibri, S.E.; Krogstie, J. Smart sustainable cities of the future: An extensive interdisciplinary literature review. *Sustain. Cities Soc.* **2017**, *31*, 183–212. [CrossRef]
5. Caragliu, A.; Bo, C.D.; Nijkamp, P. Smart Cities in Europe. *J. Urban Technol.* **2011**, *18*, 65–82. [CrossRef]
6. Hogan, A.; Blomqvist, E.; Cochez, M.; D'amato, C.; Melo, G.D.; Gutierrez, C.; Kirrane, S.; Gayo, J.E.L.; Navigli, R.; Neumaier, S.; et al. Knowledge Graphs. *ACM Comput. Surv.* **2021**, *54*, 1–37. [CrossRef]
7. Noy, N.; Gao, Y.; Jain, A.; Narayanan, A.; Patterson, A.; Taylor, J. Industry-scale knowledge graphs: Lessons and challenges. *Commun. ACM* **2019**, *62*, 36–43. [CrossRef]
8. Ji, S.; Pan, S.; Cambria, E.; Marttinen, P.; Yu, P.S. A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Trans. Neural Netw. Learn. Syst.* **2022**, *33*, 494–514. [CrossRef] [PubMed]
9. Pan, J.Z.; Vetere, G.; Gomez-Perez, J.M.; Wu, H. (Eds.) *Exploiting Linked Data and Knowledge Graphs in Large Organisations*; Springer: Berlin/Heidelberg, Germany, 2017. [CrossRef]
10. Buchgeher, G.; Gabauer, D.; Martinez-Gil, J.; Ehrlinger, L. Knowledge Graphs in Manufacturing and Production: A Systematic Literature Review. *IEEE Access* **2021**, *9*, 55537–55554. [CrossRef]
11. Harris, S.; Seaborne, A. SPARQL 1.1 Query Language. W3c Recommendation, W3C, 2013. Available online: https://www.w3.org/TR/sparql11-query/ (accessed on 26 April 2026).
12. Battle, R.; Kolas, D. *GeoSPARQL: A Geographic Query Language for RDF Data*; OGC Standard OGC 11-052r4; Open Geospatial Consortium: Arlington, VA, USA, 2012.
13. Bordes, A.; Usunier, N.; Garcia-Duran, A.; Weston, J.; Yakhnenko, O. Translating Embeddings for Modeling Multi-relational Data. In Proceedings of the Advances in Neural Information Processing Systems, Lake Tahoe, NV, USA, 5–10 December 2013; Volume 26, pp. 2787–2795.
14. Sun, Z.; Deng, Z.H.; Nie, J.Y.; Tang, J. RotatE: Knowledge Graph Embedding by Relational Rotation in Complex Space. In Proceedings of the 7th International Conference on Learning Representations (ICLR 2019), Long Beach, CA, USA, 6–9 May 2019.
15. Nickel, M.; Murphy, K.; Tresp, V.; Gabrilovich, E. A Review of Relational Machine Learning for Knowledge Graphs. *Proc. IEEE* **2016**, *104*, 11–33. [CrossRef]

16. Hamilton, W.L.; Ying, R.; Leskovec, J. Inductive Representation Learning on Large Graphs. In Proceedings of the Advances in Neural Information Processing Systems (NeurIPS 2017), Long Beach, CA, USA, 4–9 December 2017; pp. 1024–1034.
17. Wang, Z.; Han, F.; Zhao, S. A Survey on Knowledge Graph Related Research in Smart City Domain. *ACM Trans. Knowl. Discov. Data* **2024**, *18*, 1–31. [CrossRef]
18. De Nicola, A.; Villani, M.L. Smart City Ontologies and Their Applications: A Systematic Literature Review. *Sustainability* **2021**, *13*, 5578. [CrossRef]
19. Peffers, K.; Tuunanen, T.; Rothenberger, M.A.; Chatterjee, S. A Design Science Research Methodology for Information Systems Research. *J. Manag. Inf. Syst.* **2007**, *24*, 45–77. [CrossRef]
20. Gregor, S.; Hevner, A.R. Positioning and Presenting Design Science Research for Maximum Impact. *Manag. Inf. Syst. Q.* **2013**, *37*, 337–355. [CrossRef]
21. Hevner, A.R.; March, S.T.; Park, J.; Ram, S. Design Science in Information Systems Research. *Manag. Inf. Syst. Q.* **2004**, *28*, 75–106. [CrossRef]
22. Compton, M.; Barnaghi, P.; Bermudez, L.; García-Castro, R.; Corcho, O.; Cox, S.; Graybeal, J.; Hauswirth, M.; Henson, C.; Herzog, A.; et al. The SSN ontology of the W3C semantic sensor network incubator group. *J. Web Semant.* **2012**, *17*, 25–32. [CrossRef]
23. Haller, A.; Janowicz, K.; Cox, S.J.; Lefrançois, M.; Taylor, K.; Phuoc, D.L.; Lieberman, J.; García-Castro, R.; Atkinson, R.; Stadler, C. The modular SSN ontology: A joint W3C and OGC standard specifying the semantics of sensors, observations, sampling, and actuation. *Semant. Web* **2018**, *10*, 9–32. [CrossRef]
24. Lygerakis, F.; Kampelis, N.; Kolokotsa, D. Knowledge Graphs' Ontologies and Applications for Energy Efficiency in Buildings: A Review. *Energies* **2022**, *15*, 7520. [CrossRef]
25. Psyllidis, A. Ontology-Based Data Integration from Heterogeneous Urban Systems: A Knowledge Representation Framework for Smart Cities. In Proceedings of the 14th International Conference on Computers in Urban Planning and Urban Management (CUPUM 2015), Amsterdam, The Netherlands, 7–10 July 2015; pp. 1–21.
26. Bellini, P.; Benigni, M.; Billero, R.; Nesi, P.; Rauch, N. Km4City ontology building vs data harvesting and cleaning for smart-city services. *J. Vis. Lang. Comput.* **2014**, *25*, 827–839. [CrossRef]
27. Liu, Y.; Ding, J.; Fu, Y.; Li, Y. UrbanKG: An Urban Knowledge Graph System. *ACM Trans. Intell. Syst. Technol.* **2023**, *14*, 1–25. [CrossRef]
28. Wang, H.; Yu, Q.; Liu, Y.; Jin, D.; Li, Y. Spatio-Temporal Urban Knowledge Graph Enabled Mobility Prediction. *Proc. ACM Interact. Mob. Wearable Ubiquitous Technol.* **2022**, *5*, 1–24. [CrossRef]
29. Tao, F.; Zhang, H.; Liu, A.; Nee, A.Y.C. Digital Twin in Industry: State-of-the-Art. *IEEE Trans. Ind. Inform.* **2019**, *15*, 2405–2415. [CrossRef]
30. Dembski, F.; Wössner, U.; Letzgus, M.; Ruddat, M.; Yamu, C. Urban Digital Twins for Smart Cities and Citizens: The Case Study of Herrenberg, Germany. *Sustainability* **2020**, *12*, 2307. [CrossRef]
31. Curry, E. *Real-Time Linked Dataspaces: Enabling Data Ecosystems for Intelligent Systems*; Springer: Berlin/Heidelberg, Germany, 2020. [CrossRef]
32. Poveda-Villalón, M.; Fernández-Izquierdo, A.; Fernández-López, M.; García-Castro, R. LOT: An industrial oriented ontology engineering framework. *Eng. Appl. Artif. Intell.* **2022**, *111*, 104755. [CrossRef]
33. Palumbo, E.; Rizzo, G.; Troncy, R. STEM: Stacked threshold-based entity matching for knowledge base generation. *Semant. Web* **2018**, *10*, 117–137. [CrossRef]
34. Abdelaziz, I.; Harbi, R.; Khayyat, Z.; Kalnis, P. A survey and experimental comparison of distributed SPARQL engines for very large RDF data. *Proc. VLDB Endow.* **2017**, *10*, 2049–2060. [CrossRef]
35. Benaribi, F.I.; Malki, M.; Faraoun, K.M.; Ouchani, S. A SPARQL-based framework to preserve privacy of sensitive data on the semantic web. *Serv. Oriented Comput. Appl.* **2023**, *17*, 183–199. [CrossRef]
36. Hoang, A.T.; Carminati, B.; Ferrari, E. Protecting Privacy in Knowledge Graphs With Personalized Anonymization. *IEEE Trans. Dependable Secur. Comput.* **2024**, *21*, 2181–2193. [CrossRef]
37. March, S.T.; Smith, G.F. Design and natural science research on information technology. *Decis. Support Syst.* **1995**, *15*, 251–266. [CrossRef]
38. Venable, J.; Pries-Heje, J.; Baskerville, R. FEDS: A Framework for Evaluation in Design Science Research. *Eur. J. Inf. Syst.* **2016**, *25*, 77–89. [CrossRef]
39. Zhong, L.; Wu, J.; Li, Q.; Peng, H.; Wu, X. A Comprehensive Survey on Automatic Knowledge Graph Construction. *ACM Comput. Surv.* **2023**, *56*, 1–62. [CrossRef]
40. Xiao, H.; Huang, M.; Zhu, X. TransG: A Generative Model for Knowledge Graph Embedding. In Proceedings of the 54th Annual Meeting of the Association for Computational Linguistics, Berlin, Germany, 7–12 August 2016; Volume 1.
41. Pan, F.; Hobbs, J.R. Time Ontology in Owl. 2006. Available online: https://www.w3.org/TR/owl-time/ (accessed on 26 April 2026).
42. Lebo, T.; Sahoo, S.; McGuinness, D. PROV-O: The PROV Ontology. 2013. Available online: https://www.w3.org/TR/prov-o/ (accessed on 26 April 2026).

43. Bonifati, A.; Fletcher, G.; Voigt, H.; Yakovets, N. *Querying Graphs*; Synthesis Lectures on Data Management; Springer International Publishing: Berlin/Heidelberg, Germany, 2018. [CrossRef]
44. UC Irvine Machine Learning Repository. Individual Household Electric Power Consumption Data Set. Available online: https://archive.ics.uci.edu/ml/datasets/individual+household+electric+power+consumption (accessed on 26 April 2026).
45. Pollution Control Department, Thailand. Air Quality and Noise Management Bureau: Air Quality Data. Available online: http://air4thai.pcd.go.th/ (accessed on 26 April 2026).
46. Kaggle. Thailand Road Accident [2019–2022]. Available online: https://www.kaggle.com/datasets/thaweewatboy/thailand-road-accident-2019-2022. (accessed on 26 April 2026).
47. National Institute for Emergency Medicine (NIEM), Thailand. ITEMS: Information Technology for Emergency Medical Systems of Thailand. Available online: https://ws.niems.go.th/items_front/index.aspx (accessed on 26 April 2026).
48. Jientrakul, R.; Yuangyai, C.; Boonkul, K.; Chaicharoenwut, P.; Nilsang, S.; Pimsakul, S. Integrating Spatial Risk Factors with Social Media Data Analysis for an Ambulance Allocation Strategy: A Case Study in Bangkok. *Sustainability* **2022**, *14*, 10247. [CrossRef]
49. Traffy Fondue. Bangkok Citizen Issue Reporting Platform. Available online: https://traffy.in.th/ (accessed on 26 April 2026).
50. National Statistical Office of Thailand. Thailand Open Data Portal. Available online: http://www.nso.go.th/ (accessed on 26 April 2026).
51. Department of Public Works and Town & Country Planning, Thailand. Bangkok Metropolitan Land Use GIS Database. Available online: https://www.dpt.go.th/th (accessed on 26 April 2026).
52. Moreno, C.; Allam, Z.; Chabaud, D.; Gall, C.; Pratlong, F. Introducing the "15-Minute City": Sustainability, Resilience and Place Identity in Future Post-Pandemic Cities. *Smart Cities* **2021**, *4*, 93–111. [CrossRef]
53. Badii, C.; Belay, E.; Bellini, P.; Cenni, d.; Marazzini, M.; Mesiti, M.; Nesi, P.; Pantaleo, G.; Paolucci, M.; Valtolina, S.; et al. Snap4City: A Scalable IOT/IOE Platform for Developing Smart City Applications. In Proceedings of the 2018 IEEE SmartWorld, Ubiquitous Intelligence & Computing, Advanced & Trusted Computing, Scalable Computing & Communications, Cloud & Big Data Computing, Internet of People and Smart City Innovation (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Guilin, China, 8–12 October 2018; pp. 2109–2116. [CrossRef]
54. Ning, Y.; Liu, H.; Wang, H.; Zeng, Z.; Xiong, H. UUKG: Unified Urban Knowledge Graph Dataset for Urban Spatiotemporal Prediction. In *Proceedings of the Advances in Neural Information Processing Systems*; Oh, A., Naumann, T., Globerson, A., Saenko, K., Hardt, M., Levine, S., Eds.; Curran Associates, Inc.: Red Hook, NY, USA, 2023; Volume 36, pp. 62442–62456.
55. Lecue, F. On the role of knowledge graphs in explainable AI. *Semant. Web* **2020**, *11*, 41–51. [CrossRef]
56. Wang, L.; Guo, B.; Yang, Q. Smart City Development With Urban Transfer Learning. *Computer* **2018**, *51*, 32–41. [CrossRef]
57. Zhang, W.; Han, J.; Xu, Z.; Ni, H.; Liu, H.; Xiong, H. Urban Foundation Models: A Survey. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 25–29 August 2024; pp. 6633–6643. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.