Ontology-Driven Agents Skills for Selection of Building Datasets with various specifications

Ting-Yi Lee[1], Jung-Hong Hong[1]

[1]Department of Geomatics, National Cheng Kung University, Tainan, Taiwan

P66144027@gs.ncku.edu.tw, junghong@mail.ncku.edu.tw

Keywords: Geospatial dataset specifications, Building dataset selection, Ontology reasoning, Agent skills

Abstract

Selecting the correct dataset among Taiwan's nine building specifications requires knowledge of production lineage, legal basis, completeness constraints, and spatial-unit semantics that are rarely represented in machine-readable form, leaving large language model (LLM) agents prone to structurally unsuitable recommendations. This study proposes a framework that integrates OWL ontology reasoning with an agent skills architecture to address this gap. A four-dimensional attribute framework first characterizes the nine datasets across data identification, spatiotemporal characteristics, production and legal basis, and usability, separating properties suitable for formal reasoning from descriptive reference content. These formalizable properties are encoded in an OWL 2 DL ontology, where class hierarchies including a multiple-inheritance HybridData class allow a HermiT reasoner to automatically derive suitableFor and notSuitableFor relationships through subsumption, without manually authored rules. The inferred relationships are compiled into an agent skills decision tree (SKILL.md) with on-demand reference documents, letting an LLM agent return ontology-grounded recommendations for natural language queries. Evaluated against an unsupported baseline LLM on 37 direct, ambiguous, and professional queries, the ontology-grounded agent scored a perfect 10.0/10 average and won or tied every comparison, while the baseline averaged 7.5/10 and, in the sharpest case, recommended a dataset the ontology explicitly marks unsuitable. Two demonstration cases further show the framework handling both single-domain suitability rejection and cross-domain, constraint-aware integration.

1. Introduction

Geographic information systems rely on building datasets to support a wide range of analytical tasks, from spatial density estimation and coverage ratio calculation to property rights query and three-dimensional urban modeling. In practice, however, no single dataset serves all purposes. Selecting the appropriate dataset for a given task requires understanding not only what a dataset contains geometrically, but also its production lineage, legal basis, completeness constraints, and domain-specific usability conditions. This knowledge is rarely accessible in a structured, machine-readable form, and its absence creates a persistent gap between data producers and data users (Wiegand & García, 2007).

1.1 Problem statement

Taiwan presents a representative instance of this challenge. Nine building dataset specifications are currently maintained by multiple agencies including the Ministry of the Interior, land registration offices, and household registration offices, each operating under distinct production standards, legal frameworks, and spatial unit definitions (Hong et al., 2024). These datasets range from two-dimensional cadastral polygons and attribute-only registration records to LOD1 building footprints, LOD3 three-dimensional cadastral property models, and building number points that serve as cross-dataset integration keys. A dataset's identity and analytical suitability are defined not merely by its geometric boundaries but by its provenance characteristics, including the specific algorithms, standards, and institutional processes used during production (Sun et al., 2019), and such heterogeneity across multi-agency datasets is a recognized obstacle to reliable data integration more broadly (Ranatunga et al., 2025).

These semantic distinctions are non-trivial and consequential. A user querying which dataset to use for building density analysis may not recognize that building number points represent ownership units — one building may contain multiple points — rendering them structurally unsuitable for spatial extent analysis regardless of their apparent geometric relevance. The building footprint dataset, while geometrically appropriate for spatial analysis, carries no legal property information and cannot support ownership queries. The three-dimensional cadastral property model is the most capable dataset in terms of combined spatial and property semantics, yet covers only buildings registered after 2021, a completeness constraint that fundamentally limits its applicability for city-wide or historical analyses. Such distinctions require formalized, domain-specific knowledge that is currently not represented in any machine-readable specification, and whose absence forces users to rely on specialist expertise that does not travel with the data (Wiegand & García, 2007; Devillers et al., 2007).

1.2 Research gap

Large language model agents have recently shown promise for supporting geospatial data selection through natural language interaction (Pan et al., 2026). However, without sufficient domain knowledge, LLM-based approaches remain prone to semantic misinterpretation. Standalone LLMs are incapable of ensuring metric and topological consistency, producing spatial hallucinations that are functionally invalid for real-world analytical workflows, and their internal reasoning is not natively linked to structured schemas, making them unreliable where factual consistency with a specific database is required (Pan et al., 2026; Dorobantu & Badea, 2026). In specialized domains such as geospatial data retrieval, models frequently fail to capture implicit professional intents and remain limited by their own parametric knowledge boundaries when domain-specific constraints are absent (Li et al., 2025). This limitation is particularly acute in heterogeneous dataset environments where the difference between a suitable and an unsuitable recommendation depends on structural properties — such as spatial unit definition or completeness scope that are not expressed in standard metadata fields.

Ontological approaches have been proposed to formalize geospatial knowledge and support semantic interoperability. Task-based ontology research has demonstrated the value of mapping specific user activities to dataset criteria through subsumption reasoning, capturing expert knowledge to guide non-specialists toward appropriate data sources (Wiegand & García, 2007). However, this work relies on explicit declarations of data instances by providers rather than automated semantic resolution or LLM-driven natural language interaction. Other frameworks have extended ontology-based data access to integrate heterogeneous environmental datasets, but remain focused on structured relational databases without an integrated natural language interaction layer (Ranatunga et al., 2025). More recent multi-agent frameworks have grounded LLM outputs in graph-based digital twins to mitigate hallucinations, yet their evaluation remains limited to relatively simple prototypes and does not generalize to highly heterogeneous, specification-level dataset environments (Pan et al., 2026). No existing work has addressed the specific challenge of formalizing dataset selection criteria for heterogeneous building specifications while simultaneously supporting natural language interaction through an LLM agent with verifiable, inheritance-based reasoning.

A further limitation concerns scalability. Rule-based selection systems require manual maintenance and do not scale as the number of datasets or analytical contexts grows: every new dataset requires explicit authoring of all applicable rules rather than inheriting them automatically from a class hierarchy. Description Logic reasoning offers a principled alternative, as restrictions defined on a parent class are automatically and soundly propagated to all subclasses and their instances through subsumption inference (Rudolph, 2011; Grosof et al., 2003), but this capability has not been applied to the specific problem of automatically deriving and exposing dataset suitability through a natural language agent interface.

1.3 Contributions

This study proposes a framework that integrates OWL ontology reasoning with an agent skills architecture to address the limitations identified above (Figure 1). Nine Taiwan building datasets are first characterized using a four-dimensional attribute framework covering data identification, spatiotemporal characteristics, production and legal basis, and usability. The design draws on established geospatial data characterization schemes that distinguish essential, morphologic, and provenance characteristics (Sun et al., 2019) and on task-based metadata models that ground dataset selection in theme, currency, precision, and lineage (Wiegand & García, 2007), separating properties suitable for formal ontology reasoning from those better handled as natural language reference content.

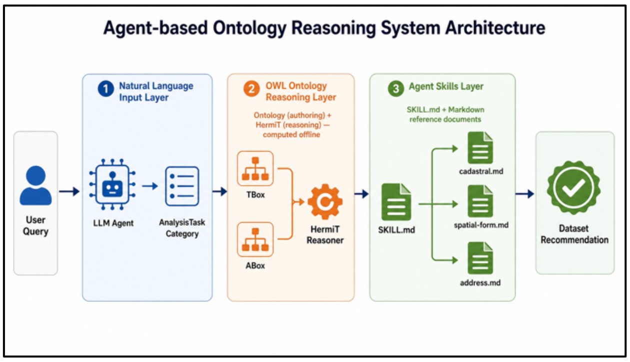

Figure 1. System architecture.

The attribute framework is formalized as an OWL ontology defining class hierarchies, object properties, and reasoning rules. A Description Logic reasoning engine automatically derives all suitableFor and notSuitableFor relationships through subclass inheritance, providing a mathematically grounded and sound basis for dataset suitability inference (McGuinness & van Harmelen, 2004; Rudolph, 2011) that eliminates the need for manually authored rules. These inferred relationships are injected into agent skills as a dataset selection decision tree, with domain-specific reference documents loaded on demand when a user query is classified into a specific domain — a design consistent with retrieval-augmented approaches that ground language model outputs in verifiable external evidence rather than relying solely on parametric knowledge (Arslan et al., 2024; Fan et al., 2024).

The framework makes three specific contributions. First, a four-dimensional attribute schema provides a structured and replicable basis for characterizing heterogeneous building dataset specifications. Second, an OWL ontology with subclass-based reasoning rules automates suitability inference, ensuring consistent derivation of suitableFor and notSuitableFor relationships across the dataset hierarchy without manual rule maintenance. Third, an ontology-augmented agent skills architecture enables LLMs to handle natural language dataset selection requests with knowledge-grounded, verifiable reasoning, combining the formal rigor of ontology inference with the flexibility of on-demand reference retrieval. The remainder of this paper is organized as follows. Section 2 reviews related work on geospatial ontologies, LLM agents for spatial data tasks, and Taiwan building dataset specifications. Section 3 describes the characterization framework, ontology design, and agent skills architecture. Section 4 presents reasoning results and two demonstration cases. Section 5 discusses contributions, limitations, and future directions. Section 6 concludes the paper.

2. Related work

2.1 Ontologies for geospatial information

OWL provides a formal, logic-based semantics that distinguishes it from informal taxonomies or plain relational schemas, enabling reasoning software to deduce that membership in a subclass necessitates satisfaction of its parent class's constraints, with OWL DL in particular ensuring the computational completeness and decidability needed for consistent, sound reasoning over a class hierarchy (McGuinness & van Harmelen, 2004; Rudolph, 2011). This formal foundation lets software reuse semantic data and reasoning rules without rewriting code, reducing the maintenance burden of rule-based alternatives (Claramunt, 2020).

Several studies have applied ontology reasoning to geospatial dataset characterization and retrieval. Sun et al. (2019) proposed GeoDataOnt, formalizing geospatial data through essential, morphologic, and provenance characteristics to establish that a dataset's identity cannot be reduced to its geometric properties alone. Wiegand and García (2007) and Klien and Probst (2005) both demonstrated task-based ontology approaches in which subsumption reasoning — rather than keyword-based queries — determines whether a dataset satisfies a specific application task, with the latter extending this to runtime semantic matchmaking so new data descriptions can be added without modifying the underlying domain logic. More recently, Ranatunga et al. (2025) extended GeoSPARQL to integrate heterogeneous environmental datasets through a virtual knowledge graph, avoiding physical data materialization while supporting both static and real-time sources.

These studies establish that OWL reasoning can automate the formal verification of dataset suitability and reduce reliance on manual rule maintenance. However, none addresses the specific structure of multiple representation building dataset specifications, where suitability depends on interacting factors such as spatial unit definition, production completeness, and legal access constraints; nor has this expert knowledge been formalized as an explicit, machine-readable suitableFor / notSuitableFor relationship that can be automatically derived through class inheritance and exposed to a natural language agent — the approach this study proposes.

2.2 LLM agents for spatial data tasks

Large language models have been increasingly applied to geospatial data tasks, but their standalone use exposes consistent domain-specific limitations: models are prone to hallucination and physically invalid recommendations because their internal reasoning is not natively linked to structured data schemas (Pan et al., 2026), struggle with procedural and quantitative reasoning such as coordinate transformations (Dorobantu & Badea, 2026), face capability boundaries in specialized fields like geosciences (Li et al., 2025), and require persistent human oversight for domain-specific edge cases such as historical naming variations (Pawełoszek, 2025).

To address these limitations, recent work has integrated LLMs with structured external knowledge. Pan et al. (2026) proposed a modular multi-agent framework grounding LLM outputs in graph-based digital twins, achieving near-perfect accuracy on building record retrieval but limited to simple prototypes. Li et al. (2025) combined a geospatial knowledge graph with spatiotemporal reasoning to let the LLM interactively explore and prune relevant facts, at high computational cost for complex queries. Liang et al. (2025) proposed a semantic-driven knowledge graph extracting modeling processes from heterogeneous geoprocessing scripts via path-based retrieval.
These approaches share a common architectural principle consistent with retrieval-augmented generation, where decoupling retrieval from reasoning improves factual accuracy while preserving interpretability (Arslan et al., 2024; Fan et al., 2024; Li et al., 2024; Liang et al., 2025). This principle extends to agent design: agent skills can be structured as reusable, trigger-able modules whose metadata is exposed for selection, with full procedural content loaded only once invoked, minimizing prompt overhead while letting effective capability scale independently of the baseline context (Ling et al., 2026; Li & Ning, 2023). However, none of the reviewed systems combines this on-demand retrieval principle with a Description Logic ontology capable of automatically deriving and propagating suitability constraints through class inheritance.

2.3 Taiwan building dataset specifications

Building datasets in Taiwan are produced under a combination of cadastral surveying regulations and topographic mapping standards administered by separate government bodies. Hong et al. (2024) provided the first systematic three-dimensional perspective on Taiwan's cadastral building data, characterizing the three-dimensional cadastral property model, which represents ownership units as LOD3 solid geometry derived from two-dimensional building survey maps — reflecting a broader international shift in which 3D cadastral models address the limitations of planar land parcels in vertically stratified developments (Aien et al., 2013; Stoter et al., 2016).

Within this context, Taiwan's three-dimensional cadastral property model occupies a distinctive position: it is the only dataset among the nine examined in this study that combines spatial form and property semantics, formally encoding ownership units, floor associations, and building registration linkages. However, consistent with the broader challenge of semantic heterogeneity in multi-agency geospatial data (Sun et al., 2019), its production lineage, legal basis, and completeness constraints — particularly its limited coverage of buildings registered after 2021 — differ substantially from the other eight datasets, reinforcing the need for a formal characterization framework.

3. Methodology

3.1 Building dataset characterization

Nine building dataset specifications were collected from three administering agencies in Taiwan: the Ministry of the Interior's

Department of Land Administration, land registration offices, and household registration offices. These datasets include the building survey map, building registration data, cadastral map, building number point, building footprint, the LOD1 three-dimensional building model, the LOD3 three-dimensional cadastral property model, the land administration address, and the household registration address.

To systematically characterize these datasets prior to ontology construction, a four-dimensional attribute framework was developed, covering data identification, spatiotemporal characteristics, production and legal basis, and usability (Table1). The four dimensions were designed to separate properties suitable for formal logical reasoning from those better expressed as natural language reference content. Dimension 1 — dataset name, dataset type, and corresponding entity — directly informs the construction of OWL classes and instances, since these properties define unambiguous class membership: dataset type determines which top-level ontology class (e.g., spatial form data, cadastral data) a dataset belongs to, and corresponding entity determines its spatial unit. Dimensions 2 through 4, by contrast, capture properties such as geometric resolution, production standards, and common use contexts that are descriptive rather than classificatory. While select properties from these dimensions — geometry type, level of detail, completeness, and access level — were formalized as object properties to support reasoning, the majority of their content (e.g., specific accuracy figures, production workflows, legal references) does not lend itself to subsumption-based inference and is instead retained as structured natural language content for on-demand retrieval.

Dimension	Field	Example — 3D Cadastral Property Model
① Data Identification		
Dim 1	Dataset name	Three-dimensional cadastral property model
Dim 1	Dataset type	HybridData (SpatialFormData ⊓ CadastralData)
Dim 1	Corresponding entity	Internal ownership unit of a building
② Spatiotemporal Characteristics		
Dim 2	Geometric dimension	3D solid
Dim 2	Spatial resolution	Inherited from BuildingSurveyMap (±2 cm / ±6 cm, urban areas)
Dim 2	Update frequency	Upon each new building registration
Dim 2	Spatial unit level	Ownership unit
③ Data Production & Legal Basis		
Dim 3	Responsible authority	Land Administration Offices
Dim 3	Reference legislation	NGISTD-ANC-042-2025 (CityGML 2.0 ADE)
Dim 3	Production process	Auto-generated from 2D building survey maps via land administration software
④ Data Usability		
Dim 4	Record content	Ownership boundary, floor height, land administration attributes
Dim 4	Completeness	Buildings registered after 2021 only
Dim 4	Data format	ZJB
Dim 4	Common use contexts	3D visualization + property rights query (cross-domain)

Table 1. Four-dimension attribute framework.

The three-dimensional cadastral property model illustrates this characterization process across all four dimensions. Under Dimension 1, it is identified as a cadastral dataset whose corresponding entity is the internal property unit of a building, leading to its formalization as a class that inherits from both spatial form and cadastral classes in the ontology. Under Dimension 2, its three-dimensional geometry, LOD3 level of detail, and cadastral survey accuracy are encoded as object property values. Under Dimension 3, its governing legislation and production standard — automatic generation from two-dimensional building survey maps via land administration software — are recorded as a data property value, since the production workflow itself is not subject to logical inference. Under Dimension 4, its completeness constraint (limited to buildings registered after 2021) is formalized as an object property to support inference, while its common use contexts are retained as natural language content for the agent skills layer.

3.2 Ontology design

The ontology was implemented in OWL 2 DL using Protégé and processed with the HermiT reasoner, chosen for its support of full OWL DL expressivity and sound, complete subsumption reasoning (Figure 2).

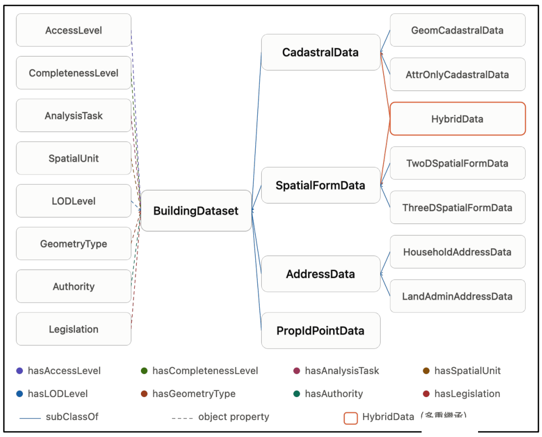

Figure 2. Tbox class hierarchy of the building dataset ontology.

Class hierarchy (TBox). The top-level class BuildingDataset has five immediate subclasses corresponding to the structural categories identified in Dimension 1: SpatialFormData, CadastralData, PropIdPointData, AddressData, and HybridData. SpatialFormData is further divided into TwoDSpatialFormData and ThreeDSpatialFormData; CadastralData is divided into GeomCadastralData (datasets with geometric representation) and AttrOnlyCadastralData (attribute-only records). HybridData is declared as a subclass of both SpatialFormData and CadastralData, formally expressing that the three-dimensional cadastral property model possesses both spatial form and cadastral semantics simultaneously, rather than belonging to a separate, disconnected category. This multiple-inheritance design allows the reasoning engine to propagate suitability restrictions from both parent branches to HybridData instances without requiring any restriction to be authored specifically for the hybrid class.

Eight supporting top-level classes were defined to serve as the range of object properties: AnalysisTask (the target of suitability reasoning, with subclasses including SpatialExtentAnalysis, PropertyRightsQuery, ThreeDVisualization, AddressGeocoding, BuildingIdentification, and CrossDatasetIntegration), GeometryType, SpatialUnit, LODLevel, CompletenessLevel, AccessLevel, Authority, and Legislation.

Object and data properties. Eleven object properties were defined with BuildingDataset as their domain, including suitableFor and notSuitableFor (range: AnalysisTask), hasGeometryType, hasSpatialUnit, hasLOD, hasCompleteness, hasAccessLevel, isManagedBy, governedBy, and integratesVia (range: BuildingDataset, declared as a symmetric property to reflect that cross-dataset integration keys, such as the building number, relate datasets bidirectionally). Six data properties — including hasDataFormat, hasSpatialAccuracy, and hasUseContext — were defined to hold descriptive string content corresponding to Dimensions 2 through 4 that does not participate in reasoning but is retrieved by the agent skills layer.

Because these restrictions are defined at the class level using OWL's subsumption semantics, any individual declared as an instance of a subclass automatically inherits the restrictions of all its ancestor classes. For HybridData, this means an instance simultaneously inherits suitableFor restrictions from SpatialFormData and CadastralData, without either restriction being explicitly re-declared on HybridData itself. This inheritance-based design directly supports the scalability goal of the framework: incorporating a new dataset into the ontology requires only declaring it as an instance of the appropriate class; all applicable suitability constraints are inferred automatically by the reasoner rather than being manually authored.

ABox. Each of the nine datasets was declared as a named individual under its corresponding class, with object property assertions populated from the formalizable subset of Dimensions 1–4 (e.g., BuildingFootprint was declared as an instance of TwoDSpatialFormData with hasGeometryType Polygon, hasLOD LOD1, and hasCompleteness Comprehensive) and data property assertions populated from the remaining descriptive content. Running the HermiT reasoner over the asserted TBox and ABox produces the complete set of inferred suitableFor and notSuitableFor relationships for all nine datasets, reported in Section 4.

3.3 Agent skills architecture

The agent skills layer translates ontology reasoning output into a form usable by a natural language LLM agent, following a structure of one trigger-loaded decision file (SKILL.md) and multiple on-demand reference documents.

SKILL.md contains a decision tree organized by AnalysisTask category. Each category entry is populated directly from the reasoner's inferred suitableFor and notSuitableFor relationships, presenting, for a given analytical task, the set of suitable datasets, the set of explicitly excluded datasets, and the dataset serving as the cross-dataset integration key where relevant. This file is loaded whenever a user query is classified as a dataset selection or suitability request, providing the agent with verifiable, ontology-grounded recommendations before any natural language elaboration is generated.

Three reference documents — cadastral.md, spatial-form.md, and address.md — correspond to the dataset groupings used in Dimension 1 and contain the non-formalizable content from Dimensions 2 through 4: production workflows, legal access conditions (such as the three-tier transcript system governing building registration data), scale-dependent specifications, and natural language use-case descriptions. These documents are not loaded by default; they are retrieved only when the decision tree identifies that the user's query concerns a specific dataset domain requiring supplementary detail beyond the formal suitability result. This on-demand retrieval design follows the principle that grounding language model outputs in external knowledge retrieved at inference time, rather than embedding all domain content in the system prompt upfront, improves factual grounding while reducing prompt overhead (Arslan et al., 2024; Fan et al., 2024). Section 4 demonstrates this architecture through two representative user queries, tracing the path from query classification through ontology-derived recommendation to reference document retrieval.

4. Result

4.1 Ontology reasoning output

Running the HermiT reasoner over the asserted TBox and ABox (Section 3.2) produces, for each of the nine datasets, the complete set of suitableFor (Table 2) and notSuitableFor (Table 3) relationships. Every entry in both tables is inferred automatically through class subsumption; none was declared directly on an individual dataset.

Dataset	suitableFor
Building footprint	SpatialExtnentAnalysis
Three dimension building model	SpatialExtnentAnalysis
	ThreeDVisualization
Three dimension cadastral model	SpatialExtnentAnalysis
	BuildingIdentification
	ThreeDVisualization
	PropertyRightsQuery
Building number point	BuildingIdentification
	CrossDatasetIntegration
Building registration data	PropertyRightsQuery
Building survey map	BuildingIdentification
	PropertyRightsQuery
Cadastral map	BuildingIdentification
	PropertyRightsQuery
Household address data	AddressGeocoding
Land administration address data	AddressGeocoding

Table 2. Inferred suitableFor relationships for the nine Taiwan building datasets, derived automatically by the HermiT reasoner from TBox subsumption rules.

https://doi.org/10.5194/isprs-archives-L-4-W2-2026-103-2026 | © Author(s) 2026. CC BY 4.0 License.

Dataset	notSuitableFor
Building footprint	ThreeDVisualization
Three dimension building model	AddressGeocoding
Building number point	SpatialExtnentAnalysis
Building registration data	SpatialExtnentAnalysis
	ThreeDVisualization
Building survey map	ThreeDVisualization
Cadastral map	ThreeDVisualization
Household address data	SpatialExtnentAnalysis
	ThreeDVisualization
Land administration address data	SpatialExtnentAnalysis
	ThreeDVisualization

Table 3. Inferred notSuitableFor relationships for the nine Taiwan building datasets.

Table 2 illustrates the inheritance mechanism described in Section 3.2. SpatialFormData is asserted suitableFor SpatialExtentAnalysis at the class level; because both BuildingFootprint (TwoDSpatialFormData) and ThreeDBuildingModel (ThreeDSpatialFormData) are declared as subclasses of SpatialFormData, both automatically inherit SpatialExtentAnalysis in the reasoner output without this relationship being asserted a second time on either dataset.

ThreeDCadastralPropertyModel illustrates the multiple-inheritance case: as a HybridData instance inheriting simultaneously from SpatialFormData and CadastralData, it is the only dataset for which the reasoner derives suitableFor across all four categories exactly the combined spatial-and-property semantics discussed in Section 2.3, obtained here without any hybrid-specific rule.

These two tables are not consulted directly by the agent at query time. Instead, they are compiled into the SKILL.md decision tree described in Section 3.3: each AnalysisTask category is populated with its corresponding suitableFor and notSuitableFor dataset lists, plus the integration key where CrossDatasetIntegration applies, so that the agent retrieves a pre-verified, ontology-grounded recommendation the moment a query is classified, before any reference document is loaded.

4.2 Case 1: Direct query

User query :Can I use building number point to calculate the building density in Tainan?

The agent skill first classifies the query as a suitability check under SpatialExtentAnalysis. Because BuildingNumberPoint is declared as an instance of PropIdPointData, for which the reasoner has already derived notSuitableFor(SpatialExtentAnalysis), the skill rejects the point dataset outright: a single physical building can carry several separately registered building numbers, one per ownership unit, so counting points would overstate density wherever multi-unit buildings exist. It then follows the SKILL.md decision-tree entry for SpatialExtentAnalysis to recommend BuildingFootprint and, where a volumetric measure is wanted, ThreeDBuildingModel, while flagging ThreeDCadastralPropertyModel only as a coverage-limited, ownership-aware alternative (Figure 3).

Not directly — BuildingNumberPoint (建號點) isn't suitable for calculating building density, even though it seems like a natural fit.

Here's why: BuildingNumberPoint represents property ownership units, not spatial extents. One physical building can correspond to multiple points — for example, an apartment building with 50 units would have 50 building number points, even though it's a single structure. So counting points would badly overestimate density, and the count doesn't tell you anything about footprint area or spatial coverage.

For density calculations in Tainan, you'd want one of these instead:

- **BuildingFootprint** — polygon geometry, public access, available at two scales (1/1,000 per-building resolution or 1/5,000 district-level). This is the standard choice for coverage ratio (建蔽率) and density work.
- **ThreeDBuildingModel** — if you also want height/volume, this gives LOD1 3D solids with comprehensive coverage.
- **ThreeDCadastralPropertyModel** — only if you need ownership semantics alongside spatial extent, but note it's post-2021 buildings only, so not suitable for city-wide or historical density analysis.

If your actual goal is something like "density of ownership units" (e.g., housing units per area) rather than building footprint density, BuildingNumberPoint could work for that different metric — but for true building/footprint density in Tainan, go with BuildingFootprint.

Figure 3. Case1 Agent response.

4.3 Case 2: Cross-domain query

User query:How can I calculate the ratio of building and ownership unit in 2000 Taiwan?

This query spans two AnalysisTask categories at once, triggering the CrossDatasetIntegration entry in SKILL.md: the building count requires BuildingFootprint (SpatialExtentAnalysis), while the ownership-unit count requires BuildingNumberPoint or BuildingRegistrationData (PropertyRightsQuery), joined through the building number as the shared key. Beyond dataset pairing, the response also surfaces a constraint outside the suitableFor / notSuitableFor relation itself: none of the nine datasets is a historical time series, so a year-2000 figure cannot be obtained from a current extract. The skill accordingly redirects the query to BuildingRegistrationData's completion-date attribute, filtering records with completion date < 2000 to reconstruct both counts for that year, and flags that bulk access to this attribute is governed by the three-tier transcript system rather than open download (Figure 4).

For this you need two counts: **number of physical buildings** and **number of ownership units** for the same area, then divide.

Metric	Dataset to use	Why
Building count	BuildingFootprint	Polygon per physical building, comprehensive coverage — use 1/1,000 scale for per-building precision
Ownership unit count	BuildingNumberPoint or BuildingRegistrationData	Each point/record = one registered ownership unit; a single building can have many (e.g. condo units)

Ratio = building count ÷ ownership unit count (per area, e.g. per village/district or grid cell). A ratio well below 1 indicates multi-unit buildings (apartments, condos) dominate; closer to 1 indicates mostly single-ownership structures (townhouse-style buildings).

The year-2000 catch: none of these datasets are historical time-series — they represent *current* inventories, updated periodically (BuildingFootprint roughly every 2 years; BuildingNumberPoint on each new registration event). So you can't just download "the 2000 snapshot." To reconstruct 2000-era figures you have two options:

1. **Filter by date field**, if available:
   - BuildingRegistrationData carries a completion/registration date in its building description section — filter records with completion date ≤ 2000 to get ownership units existing by 2000.
   - BuildingFootprint has no construction-date attribute of its own, so building age isn't directly filterable from the footprint layer.
2. **Cross-reference via building number → parcel number → address** to link BuildingRegistrationData's completion date to the corresponding building footprint, effectively backdating which footprints existed in 2000.

Given this, a cleaner approach: use **BuildingRegistrationData filtered to completion date ≤ 2000** to count both the number of distinct buildings (unique addresses/parcels, since its building description section records building-level info too) and the number of ownership units (count of building-number records) — since it's the only dataset in this set that natively carries a historical date field.

One practical constraint: BuildingRegistrationData is attribute-only (no geometry) and partially public (three-tier transcript system applies), so bulk historical queries may require going through the appropriate transcript-access tier rather than open download. Want me to sketch out the actual join/query logic (e.g., in SQL or Python) once you know which access tier you can use?

Figure 4. Case2 Agent response.

5. Discussion

5.1 Contributions of inheritance-based reasoning

HybridData's multiple inheritance (Sections 3.2, 4.1) is the central technical contribution of this framework, demonstrating that the scalability goal stated in Section 1 holds in practice:

incorporating a tenth dataset requires only declaring its class membership, after which all suitableFor / notSuitableFor relationships follow automatically through subsumption, with no manually authored rule and no risk of inconsistent hand-written exceptions as the dataset inventory grows.

5.2 Comparison with unsupported LLM recommendations

To test whether ontology-grounded reasoning yields practically better recommendations than an LLM operating without domain-specific knowledge, the agent skills system was compared against an unsupported baseline model (the same underlying LLM with web search enabled, but no ontology or skill files) on the 37-question evaluation set. Each answer was scored against the ontology-derived ground truth on five criteria — task classification, dataset selection accuracy, exclusion correctness, reasoning groundedness, and constraint awareness — following a rubric in which each criterion is worth up to two points (Table 4).

The skills system achieved a perfect 10.0/10 average across all 37 questions, while the unsupported baseline averaged 7.5/10, winning outright on none of the 37 questions and tying only on 16 (43%), chiefly on questions about publicly documented legal or access procedures that do not depend on dataset-suitability semantics. The baseline's weakest performance clustered on Direct suitability questions, where it frequently reconstructed the correct dataset category from general GIS knowledge but recommended or failed to exclude datasets that the ontology marks notSuitableFor.

The starkest case was Q11 (“Is building point data suitable for building density mapping?”): the baseline model concluded the point dataset was suitable, directly contradicting the ontology's inferred notSuitableFor because it had no structural basis for recognizing that a building number point represents an ownership unit rather than physical extent. This single case, scoring 1/10 for the baseline against 10/10 for the skills system, illustrates that the absence of formalized structural semantics rather than any general geospatial knowledge gap is what drives unsuitable recommendations in the highest-risk category of query.

Category	Skills avg (/10)	Baseline avg (/10)	Skills wins
Direct (n=12)	10	6.1	8
Ambiguous (n=12)	10	7.6	6
Professional (n=13)	10	8.8	7
Overall (n=37)	**10**	**7.5**	**21**

Table 4. Scores by query category.

5.3 Limitations and future work

This study has three limitations that motivate future work. First, the ontology and ABox are limited to nine Taiwan-specific building datasets characterized manually by domain experts; extending coverage to other jurisdictions would require repeating this characterization process, which is not yet automated from source metadata catalogs. Second, the 37-question evaluation set was authored by the same team that built the ontology, so independent user studies with practitioners unfamiliar with the framework are needed to confirm generalizability. Third, and most importantly, the framework as demonstrated in Section 4 stops at recommending which dataset (and, for Case 2, which attribute and filter) to use, it does not yet execute the query against real data. Closing this gap requires extending the ontology below the dataset level to the semantic-object level, this means formalizing sub-building components as their own OWL classes, consistent with the nested-class hierarchy already identified in Section 3.1, so that a query resolves to a specific story or unit rather than to the dataset as a whole; for BIM-sourced content, this means mapping IFC entity types (e.g., IfcSpace, IfcBuildingStory, IfcElement) into the same class hierarchy so BIM data can be reasoned over with the same suitableFor / notSuitableFor logic as the nine base datasets rather than treated as an opaque file. Once the reasoner identifies not only the suitable dataset but the semantic object and constraint within it, that recommendation needs to be compiled into an executable SQL query against the underlying relational or spatial database, so the agent skill returns actual attribute values, counts, or geometries rather than terminating at a dataset name. Realizing this SQL-execution layer would convert the framework from a dataset-selection advisor into a complete natural-language-to-data query pipeline, and is the direction of our ongoing work.

6. Conclusion

This paper presented an ontology-augmented agent skills framework for selecting among Taiwan's nine multiple representation building dataset specifications. A four-dimensional attribute schema separated formalizable properties — dataset type, geometry, spatial unit, completeness, access level — from descriptive reference content, supporting an OWL 2 DL ontology in which suitableFor and notSuitableFor relationships are derived automatically through class subsumption rather than authored per dataset. These inferred relationships are injected into a SKILL.md decision tree that an LLM agent loads on demand, with three reference documents supplying legal, scale, and integration-key detail beyond the formal reasoning result.

Evaluated against an unsupported baseline LLM on 37 direct, ambiguous, and professional queries, the ontology-grounded agent scored a perfect 10.0/10 average and won or tied every comparison, while the baseline averaged 7.5/10 and, in the sharpest case, recommended a dataset the ontology explicitly marks unsuitable. This confirms that formalizing structural dataset semantics, rather than relying on an LLM's general geospatial knowledge, is what prevents such misclassifications. The two demonstration cases further show the framework correctly handling both single-domain suitability rejection and cross-domain, constraint-aware integration. Closing the remaining gap between recommending a dataset and returning its actual data — by extending the ontology to semantic level building components and coupling reasoning output to executable SQL queries is the next step toward a complete natural-language-to-data pipeline for building datasets.

Reference

Aien, A., Kalantari, M., Rajabifard, A., Williamson, I., Wallace, J., 2013. Towards integration of 3D legal and physical objects in cadastral data models. Land Use Policy, 35, 140-154. https://doi.org/10.1016/j.landusepol.2013.05.014

Arslan, M., Ghanem, H., Munawar, S., Cruz, C., 2024. A survey on RAG with LLMs. Procedia Computer Science, 246, 3781-3790. https://doi.org/10.1016/j.procs.2024.09.178

Claramunt, C., 2020. Geospatial ontologies: A vision paper. Journal of Spatial Information Science, 2020(20), 1-7.

Devillers, R., Bédard, Y., Jeansoulin, R., Moulin, B., 2007. Towards spatial data quality information analysis tools for experts assessing the fitness for use of spatial data. International Journal of Geographical Information Science, 21(3), 261-282. https://doi.org/10.1080/13658810600912327

Dorobantu, G.I., Badea, A.C., 2026. Geospatial reasoning and awareness in large language models: A systematic review. Artificial Intelligence Review, 59, 111. https://doi.org/10.1007/s10462-026-11512-x

Fan, W., Ding, Y., Ning, L., Wang, S., Li, H., Yin, D., Chua, T.S., Li, Q., 2024. A survey on RAG meeting LLMs: Towards retrieval-augmented large language models. In Proceedings of the 30th ACM SIGKDD Conference on Knowledge Discovery and Data Mining (KDD '24) (11 pages). https://doi.org/10.1145/3637528.3671470

Grosof, B.N., Horrocks, I., Volz, R., Decker, S., 2003. Description Logic Programs: Combining logic programs with Description Logic. In Proceedings of the 12th International Conference on World Wide Web (pp. 48-57). https://doi.org/10.1145/775152.775160

Hong, J.H., Chiang, V., Yang, C.S., Tsai, Y.C., 2024. 3D perspective towards the modelling and applications of cadastral building data in Taiwan. In Proceedings of the 12th International FIG Workshop on the Land Administration Domain Model & 3D Land Administration, Kuching, Malaysia, 24-26 September 2024.

Klien, E., Probst, F., 2005. Requirements for geospatial ontology engineering. In Proceedings of the 8th AGILE Conference on Geographic Information Science (pp. 11-14).

Li, H., Yue, P., Wu, H., Teng, B., Zhao, Y., Liu, C., 2025. A question-answering framework for geospatial data retrieval enhanced by a knowledge graph and large language models. International Journal of Digital Earth, 18(1). https://doi.org/10.1080/17538947.2025.2353122

Li, Y., Zhang, R., Liu, J., 2024. An enhanced prompt-based LLM reasoning scheme via knowledge graph-integrated collaboration. In M. Wand et al. (Eds.), Artificial Neural Networks and Machine Learning – ICANN 2024 (pp. 251-265). Springer. https://doi.org/10.1007/978-3-031-72344-5_17

Li, Z., Ning, H., 2023. Autonomous GIS: The next-generation AI-powered GIS. International Journal of Digital Earth, 16(2), 4668-4686. https://doi.org/10.1080/17538947.2023.2278895

Liang, J., Hou, S., Zhao, A., Xu, Q., Xiang, L., Li, R., Wu, H., 2025. Design and application of a semantic-driven geospatial modeling knowledge graph based on large language models. Geo-spatial Information Science. https://doi.org/10.1080/10095020.2025.2497810

Ling, G., Zhong, S., Huang, R., 2026. Agent skills: A data-driven analysis of Claude skills for extending large language model functionality. arXiv preprint. arXiv:2602.08004v1

McGuinness, D.L., van Harmelen, F., 2004. OWL Web Ontology Language overview (W3C Recommendation). World Wide Web Consortium. https://www.w3.org/TR/2004/REC-owl-features-20040210/

Pan, Y., Wang, M., Lu, L., Lamsal, R., Pärn, E., Zlatanova, S., Brilakis, I., 2026. LLM-enabled multi-agent framework for natural language interaction with graph-based digital twins. Automation in Construction, 183, 106791. https://doi.org/10.1016/j.autcon.2026.106791

Pawełoszek, I., 2025. Leveraging generative AI for spatial data preparation. Procedia Computer Science, 270, 2356-2365. https://doi.org/10.1016/j.procs.2025.09.357

Ranatunga, S., Ødegård, R.S., Jetlund, K., Onstein, E., 2025. Use of Semantic Web technologies to enhance the integration and interoperability of environmental geospatial data: A framework based on ontology-based data access. ISPRS International Journal of Geo-Information, 14(2), 52. https://doi.org/10.3390/ijgi14020052

Rudolph, S., 2011. Foundations of Description Logics. In A. Polleres, C. d'Amato, M. Arenas, S. Handschuh, P. Hitzler, G.I.J.Z. Pan (Eds.), Reasoning Web. Semantic Technologies for the Web of Data (pp. 76-136). Springer. https://doi.org/10.1007/978-3-642-23032-5_2

Stoter, J., Ploeger, H., Roes, R., van der Riet, E., Biljecki, F., Ledoux, H., 2016. First 3D cadastral registration of multi-level ownership rights in the Netherlands. In Proceedings of the 5th International FIG 3D Cadastre Workshop (pp. 491-504).

Sun, K., Zhu, Y., Pan, P., Hou, Z., Wang, D., Weirong, L., Song, J., 2019. Geospatial data ontology: The semantic foundation of geospatial data integration and sharing. Big Earth Data, 3(3), 269-296. https://doi.org/10.1080/20964471.2019.1661662

Wiegand, N., García, C., 2007. A task-based ontology approach to automate geospatial data retrieval. Transactions in GIS, 11(3), 355-376. https://doi.org/10.1111/j.1467-9671.2007.01050.x