Construction and Updating Technology for a Spatiotemporal Evolution Knowledge Graph of Geographic Entities: Using National Fundamental Geographic Entity Data in China as Study Case

Shiquan Zhao[1], Jianjun Liu[1], Jianwei Liu[1], Yin Gao[1], Bianli Zhao[1], Hao Li[1], Jian Che[1], Xue He[1], Zhaokun Zhai[1]

[1] National Geomatics Center of China, 100830, Beijing, China - zhaosq@ngcc.cn, liujj@ngcc.cn, liujw@ngcc.cn, gaoyin@ngcc.cn, zbl@ngcc.cn, lihao@ngcc.cn, chejian@ngcc.cn, Hex@ngcc.cn, zhaizhaokun@ngcc.cn

Keywords: Geographic entity, Knowledge graph, Spatiotemporal process simulation, Full lifecycle management, 3D real scene

Abstract

A geographic entity is an abstract representation of a real-world geographic object or phenomenon, characterized by its location, extent, processes, interrelationships, and attributes. It serves as a fundamental unit for spatial process modelling across various domains, such as land cover change simulation, wetland management, geological monitoring, etc. Yet, current research typically emphasizes either definitional frameworks for entities or domain-specific knowledge graphs tailored to one-off analytical use cases in specific fields, thereby constraining generalizability and impeding sustained applicability in the absence of dynamic update mechanisms. To address these challenges and enable full lifecycle management and spatiotemporal knowledge discovery for geographic entities, this study introduces the Spatiotemporal Evolution Knowledge Graph (SEKG) along with its incremental update methodologies, a generalized framework built on the Property Graph model and derived from analyses of how geographic entities and their interrelationships evolve in real-world. Then, we validated the SEKG's robustness and effectiveness by implementing it based on subset of China's National Fundamental Geographic Entity Data using Neo4j and by conducting a series of comparative experiments against PostGIS-based geographic entity database. The validation demonstrates that SEKG could effectively and cohesively organize comprehensive geographic entity information, including location, extent, attributes, interrelationships, and evolution processes, while supporting incremental updates for long-term lifecycle management of geographic entities. Furthermore, as a semantic-based spatiotemporal knowledge structure, the SEKG could facilitate advanced applications such as spatiotemporal evolution simulation, semantic mining, etc. It is expected that this work may serve as a valuable reference for entity-based data management, spatiotemporal modelling, and knowledge discovery.

1. Introduction

A geographic entity is an abstract representation of a real-world geographic object or phenomenon, characterized by its location, extent, processes, interrelationships, and attributes (Bateman and Farrar, 2004). It serves as a fundamental unit for spatial process modelling across various domains, such as land cover change simulation, wetland management, geological monitoring, etc. (Wang et al., 2019; Zhang et al., 2024; Zheng et al., 2021).

In related research, developing a unified framework capable of managing and updating the multidimensional spatiotemporal characteristics of geographic entities is regarded as a fundamental requirement for spatiotemporal simulation and digital twin construction (Shen, 2018). However, most research typically emphasizes either definitional frameworks for entities or domain-specific knowledge graph construction tailored to one-off analytical use for cases in specific fields. Few studies have addressed the need for a general framework capable of organizing multidimensional geographic information comprehensively, including location, extent, attributes, spatiotemporal processes and interrelationships, while also incorporating a mechanism for updating the knowledge graph. This hinders their ability to support generalized supplication scenarios and longer-term entity-based management, simulation and data mining (Xue et al., 2019).

Currently, the prevailing technology paradigms for organizing spatiotemporal data of entities resides in tabular database architectures, such as relational databases (e.g., Oracle, PostGIS) or proprietary spatial databases (e.g., File Geodatabase). These databases organize entity attributes and spatial information as records, fundamental unit in tables, using discrete tables with different schemas, geometric types (e.g., point, polyline, polygon), and temporal versions. Leveraging the structured tabular storage model, these databases achieve high efficiency in retrieving entity records through optimized indexing mechanisms and normalized data organization (Breunig et al., 2020). However, this structured data model inherently limits the use of semantic information due to its discrete, table-based storage. Entities and relationships with different attribute structures or version statuses must be stored in separate tables. When performing complex semantic queries, such as deep relationship retrieval, associated entity statistics, or holistic provenance tracing of entity evolution, it compels repeated conditional query execution across schema-disparate tables to assemble composite results. This imposes significant computational penalties for heterogeneous entity retrieval, while concurrently constraining spatiotemporal semantic analysis and mining capabilities with complicated semantic query logic in developing related application based on such a database.

To address the limitations of tabular database in semantic retrieving and mining, graph-based architectures have been widely used in related works (Arvor et al., 2013; Ji et al., 2022; Shi et al., 2024). Generally, researchers convert spatiotemporal information into semantic descriptions and constructs knowledge graphs using graph databases (Abu-Salih, 2021; Ma, 2022). This approach liberates the same type of data (such as attribute) from the rigid constraints of table structures and efficiently associates entity nodes and relationship edges through physical pointers and link table, enabling high-performance semantic information organization and retrieval. Simultaneously, as a semantic-oriented data organization model, it supports various semantic processing tools, thereby better facilitating data mining and knowledge discovery (Lawley et al., 2023; Mai et al., 2020; Zhu et al., 2025).

However, related research on geographic knowledge graphs exhibits two main shortcomings. On the one hand, most studies represent entity spatial locations and extents using simplified positioning points. Few studies have explored graph-based organization and management for comprehensive spatial information of geographic entities, particularly when multiple geometric types are simultaneously applied to represent different spatial patterns and characteristics for one entity with respect to different application scenarios, such as using polyline to represent orientation of the entity and using polygon to represent spatial extent of the entity. On the other hand, inadequate research proposed incremental updating techniques for graphs. While some studies have proposed knowledge graph for spatiotemporal process simulation, efforts often focus on one-off use for specific analytical purposes (Ma, 2022; Xue et al., 2019). Approaches for long-term updating these knowledge graphs remain largely unaddressed. However, entities and inter-entity relationships in the real-world undergo changes over time. When building a knowledge graph without corresponding incremental update technology, it won't keep pace with the changes in the real world. This makes it hard to support long-term application of related knowledge graphs in related fields.

To address the above-mentioned challenges, this study proposed the Spatiotemporal Evolution Knowledge Graph (SEKG) along with its incremental update methodology. Firstly, we analyzed different scenarios of spatiotemporal evolution process of geographic entities and inter-entity relationships and designed corresponding representation models. Subsequently, the overall structure of the knowledge graph was constructed, along with its incremental updating methodology designed with respect to current Geographic Entity Production technique regulations used in China. Based on this, a SEKG were implemented based on a subset of National Fundamental Geographic Entity data in China over the past three years using Neo4j graph database to validate the feasibility of the proposed framework for full lifecycle management and long-term use of geographic entity and inter-entity relationships. Finally, to evaluate the strengths and limitations of SEKG in geographic entity data management and retrieval, a relational database was also built using PostGIS based on the same data as a control group. Based on this, five semantic retrieval scenarios and two spatial retrieval scenarios were designed as benchmark samples to discuss the capability and efficiency of SEKG with res PostGIS-based geographic entity data. With this, it is expected to provide technical support for full lifecycle management of geographic entities for the construction of 3D Real Scene China and to provide technical references for studies on entity-based spatiotemporal modeling, simulation and semantic mining.

2. Data source

The data basis of this study is a subset of National Fundamental Geographic Entity data products in China through the years of 2022 to 2024. The data was derived from the 1:50000 Fundamental Geographic Information data based on corresponding data production technique requirements that define the classification framework of geographic entities, incremental update data production methods and data model for geographic entity data organizations based on the technical requirements for New Fundamental Surveying and Mapping and 3D Real Scene China Construction launched by the Ministry of Natural Resources of the Peoples' Republic of China (Xiao et al., 2021; Zhang and Tian, 2025; Zheng et al., 2023).

According to the above-mentioned technical regulations, all entities in the Fundamental Geographic Entity data product are categorized into three types: Natural Geographic Entity, Artificial Geographic Entity and Management Geographic Entity. The three types are further branches into 15 level-2 subcategories and over 100 level-3 subcategories (Zheng et al., 2023). The data is stored in File Geodatabase (*.gdb) format and organized by data layers and attribute fields in accordance with the technical specifications for 3D Real Scene China Construction. The data captured in different years are distinguished by annual version labels of tables, comprising approximately 50 million entities and 5 million relationships triplets and covering an area of about 4.9 million square kilometers across multiple provinces such as Gansu, Qinghai, Shaanxi, Heilongjiang, and Shenyang etc. The data includes spatial, attribute, and relationship information of entities. Spatial data represents entities using 2D points, lines, and polygons to describe their spatial positions and extents. Some entities have multiple types of geometric representations (e.g., using both polyline and polygon to represent rivers), identified by the same unique ID and stored in separate tables corresponding to their geometric type. Attribute data includes fundamental information of entities, such as the entity's unique ID, entity class code, historical entity unique ID and temporal attributes (born time, update time, etc.), as well as thematic attributes relevant to different types of geographic entities (e.g., road numbers for road entities, hydrographic codes for river entities). Each geographic entity is distinguished by its unique ID. Relationship data captures spatial, affiliation or compositional relationships between different entities of interest, such as adjacent to relationship between peer-level administrative regions, belong to relationships between higher- and lower-level administrative divisions, pass through relationships between roads and administrative regions, flow through relationships between water bodies and administrative areas, connection relationships between different roads, flow into relationships between rivers and reservoirs or major lakes, and affiliate to relationships between ancillary facilities and major roads or water bodies.

3. Analysis on evolutionary scenarios of geographic entities and inter-entity relationships

The persistence and evolution of entities and inter-entity relationships is a continuous process. However, continuously recording such processes when digitalizing entity for management would lead to massive data redundancy. Therefore, the continuous processes were usually discretized, focusing on different states of entity and interrelationships at different time points. Accordingly, this study analyses the evolution of entities and relationships primarily based on their states at two time points along with the following core analytical criteria: (1) entity persistence is determined by the continued existence of unique entity ID; (2) an entity continues to exist during while it develops only changes its attributes and spatial morphology, while maintaining a constant unique ID; (3) inter-entity relationship is determined based on related entity IDs and relation type between the entities, any element of the triplet changes indicates the develop of the inter-entity relationship. Based on this, this study explores different evolutionary scenarios of entities and their interrelationships and designs a representation model on property graphs to support the construction and updating of SEKG.

3.1 Evolutionary scenarios of geographic entities and corresponding representation models

Generally, the evolutionary process of geographic entities can be categorized into four types: formation, development, cessation, and no change. Formation marks the beginning of an entity's lifecycle, where the entity does not exist at the previous time but emerges at the subsequent time. Development refers to the entity undergoing changes between two time points where its primary characteristics remain unchanged, involving only alterations in non-critical attributes or minor spatial geometric changes. Cessation signifies the end of the entity's lifecycle, where its unique ID ceases to exist at the subsequent time point. No change indicates that the entity's state remains completely consistent between two times. Furthermore, considering whether an entity's change involves interactions with other entities, each change type can be further categorized into two fundamental scenarios: independent evolution and joint evolution. Based on this, the evolutionary process of entities can be categorized into ten scenarios shown as Figure 1.

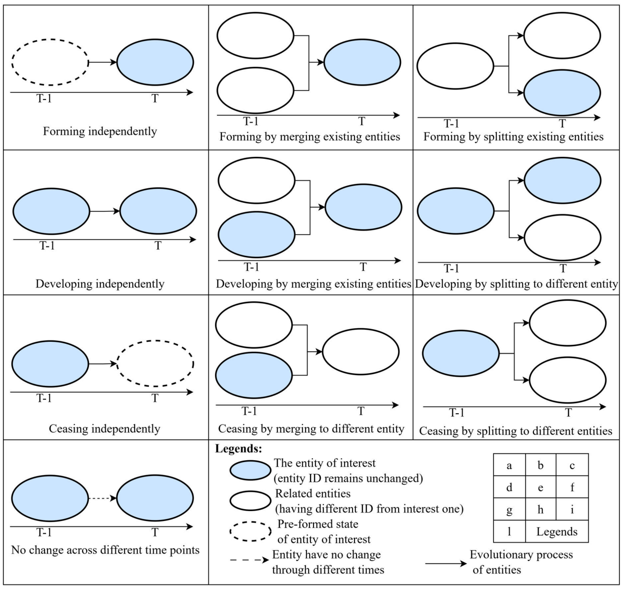

Figure 1 Evolutionary scenarios of geographic entities

In Figure 1, (a) to (c) are scenarios of entity formation. (d) to (f) are scenarios of entity development. (g) to (i) are scenarios of entity cessation. (l) is the scenario that entity has no change. (a), (d), (g) and (l) are scenarios that entity evolves independently while others are scenarios representing that entity evolves jointly with other related entities with processes like merging or splitting.

To concisely represent the evolutionary scenarios, this study focuses on the evolution of targeted entities and delineates corresponding represents models based on a property graph. In the graph nodes denote different entity states and edges denote evolutionary process. An unchanged entity retains its current state node while a changed entity (via formation, development, or demise) is captured by four evolution processes: *formation*, *update*, *succession*, or *cessation*, as detailed below.

Formation: The fundamental form for generating new entities, describing the evolution scenario where an entity is independently formed. In the property graph, only the resulting entity state node is constructed, recording information such as the generation time and storage time of the new entity, as shown in Figure 2-a.

Update: The unique entity ID remains unchanged, while some attribute information or spatial information undergoes minor changes. It primarily describes two scenarios: independent development and joint development. In the property graph, a new state node is created for the entity, recording its updated spatial and attribute information, along with the update timestamp. An *update* relationship is established between this new node and the previous state node, as shown in Figure 2-b、Figure 2-d、Figure 2-f.

Cessation: The terminal state of an entity's lifecycle. It primarily describes scenarios of independent cessation or where causally related changes lead to an entity's cessation. In the property graph, a cessation state node is constructed for the entity, recording its time of cessation. A *cessation* relationship is established between this node and the previous state node, as shown in Figure 2-c、Figure 2-e、Figure 2-g.

Succession: Describes the process of fission (splitting) or fusion (merging) between existing entities and associated entities, accompanied by the demise or renewal of existing entities, and in some cases, the generation of new entities. In the property graph, new state nodes are created for both the entities formed through succession and the original entities. The new entity state node records comprehensive information such as the generation time and historical identity codes and establishes an evolutionary *succession* relationship with the original entity. Examples of such cases are illustrated in Figure 2-d through Figure 2-g.

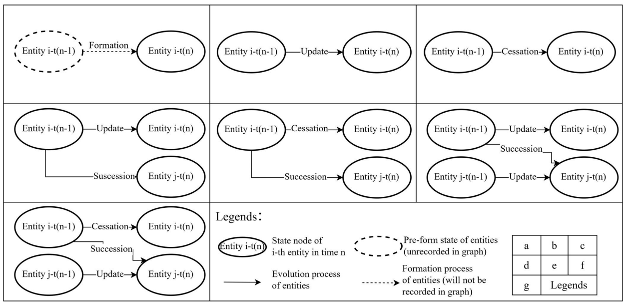

Figure 2 Representation model for evolutionary process of geographic entity based on property graph

3.2 Evolutionary scenarios of inter-entity relationship and corresponding representation models

The evolution of relationships can be categorized into three scenarios: *formation*, *discontinuation*, and *no change*. *Formation* marks the beginning of a relationship's lifecycle, where the relationship did not exist at a prior time point but emerges at a subsequent time point, with the corresponding triple first appearing in the database. *Discontinuation* signifies the end of a relationship's lifecycle, indicating that the relevant entities no longer maintain that specific type of association. *No change* means the state of the relationship remains entirely consistent between two time points. However, since relationships function as predicates connecting entities in semantic triples, it is essential to simultaneously consider whether related entity nodes within the triple have changed. This leads to two types of evolution: independent relationship evolution and joint evolution between entities and relationships. In total, there are nine fundamental scenarios, as illustrated in Figure 3.

In Figure 3, (a) to (c) are scenarios that inter-entity relationship forms through different times. (a) represents the relationship forms independently while (b) and (c) are scenarios that the relationship forms accompanied with related entity evolution. (d) to (f) are scenarios that inter-entity relationships discontinue through different times. (d) represents the relationship discontinues independently while (e) and (f) are scenarios that the relationship discontinues accompanied with related entity evolution. (g) to (i) are scenarios that inter-entity relationships persist through different times. (g) represents that both relationship and entities experienced no changes while (h) and (i) are scenarios that the relationship persists accompanied with related entity evolves (only for update and cessation scenarios).

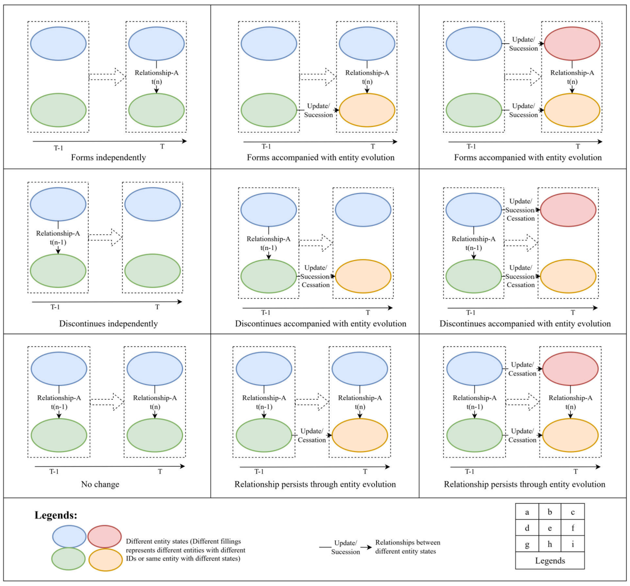

Figure 3 Evolution scenarios of correlative relationships between entities

To concisely represent the evolutionary scenarios, this study focuses on the triplet of a relationship and employs three modes, *formation*, *discontinuation*, and *persistence*, to describe the relationship evolution scenarios. Besides, temporal information of the relationships are also recorded through three attributes assigned to relationship edges: formation time, load time, and termination time.

Formation: The fundamental form of relationship emergence between entities, describing the *formation* process of a relationship. In the property graph, this type of relationship is constructed between the latest state nodes of the two entities, with the "generation time" and "archival time" assigned to the current time node. Partial case examples are shown in Figure 4-a、Figure 4-b.

Discontinuation: The point where this type of relationship between entities ceases to exist, describing the *discontinuation* process of the relationship. In the property graph, this is recorded for the corresponding relationship within the triple, with the current time marked as the termination time. Partial case examples are shown in Figure 4-d, Figure 4-e.

Persistence: The state where this type of relationship between entities persists, describing the *no change* scenario. In the property graph, there are two different scenarios that need to be discussed separately. (1) All entity states and interrelationships remain unchanged: As the triple containing the relationship undergoes no changes, no action is required for the relationship itself. (2) Entity changes while their interrelationships remain unchanged: Due to the emergence of new state nodes for the associated entities, this type of relationship needs to be constructed between the updated or succeeded entity state nodes. The "generation time" of the original relationship is retained, and the "archival time" is assigned to the current time node. Partial case example is shown in Figure 4 -c.

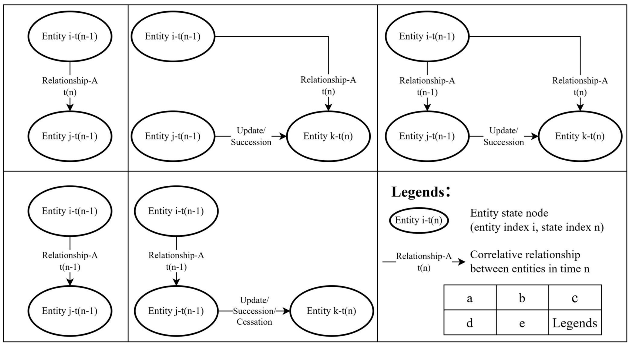

Figure 4 Representation model for evolutionary process of interrelationships based on property graph

4. Construction of SEKG

4.1 Concept model design

This study proposes the SEKG to organize comprehensive information of geographic entities and interrelationships with the framework of graph knowledge. The SEKG categorizes all types of geographic entity data into three types, spatial data, attribute data, and relationship data, to represent geographic entities, interrelationships and their evolutionary processes. The concept model is illustrated as Figure 5.

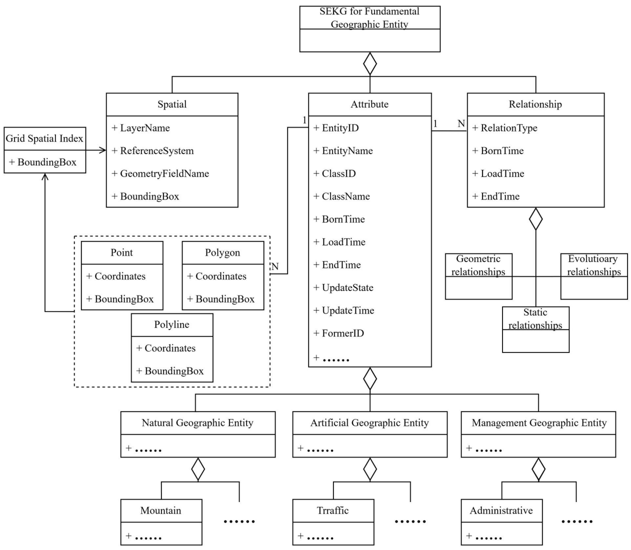

Figure 5 Concept model for SEKG

As is shown in the concept model, the spatial data consists of two-dimensional geometries, including points, polylines and polygons, to represent the spatial location and extents of the geographic entities. The geometries are restored through a sequence of coordinates within specific coordinate systems in Well-Known Text (WKT) format. Meanwhile, the bounding box of the geometry is also stored to optimize spatial retrieval performance. To manage geometric data across both entity and layer scales, this study maintains distinct spatial layers that store the entity in the original data product with corresponding grid-based spatial index for efficient spatial querying.

The attribute data of the entities mainly includes common attributes, temporal attributes, update identifier attributes and thematic attributes. The common attributes store fundamental information of entities, including entity ID, which serves as the unique identifier distinguishing entities, entity class information (entity class code and class name) and entity name. Temporal attributes provide time-series signatures of entities, including born time, end time and load time to mark the time that an entity was formed, ceased, and loaded into the graph. Thematic attributes denote specific information of different types of entities, which is different over the entity classes. Generally, the thematic attributes are designed with respect to the application scenario of geographic entity data, which have been elaborated in related studies and corresponding Technical Specification for Data Production (Zhang and Tian, 2025; Zheng et al., 2023).

The relationship data in the SEKG are categorized into three types: evolutionary relationships, geometric relationships and inter-entity relationships. The evolutionary relationships are designed to represent the evolution process between different entity states, which includes the type of *update*, *succession* and *cessation*. The geometric relationships are designed to represent the relationship between geometric representation and semantic representation of the entities. As there are three types of geometric representations for entities, there are three types of geometric relationships: *Has polyline*, *Has point* and *Has polygon*. The inter-entity relationships are triplets in the original geographic entity products to represent spatial relationships (e.g., *neighbor to*, *flow through*, *pass through*), affiliation relationships (e.g., *belong to*, *affiliate to*), composite relationships (e.g., *composition*), etc. In this study, the type of inter-entity relationships is determined based on the data source, which has been elaborated in section 2. Besides the type of relationships, the relationship data also includes temporal attributes to provide time-series signatures of the relation between corresponding entities, including born time, end time and load time to mark the time that the relation was formed, discontinued, and loaded into the graph.

4.2 Structure design for SEKG

This research employs property graphs to construct the SEKG. Entity attributes, spatial information, evolutionary relationships and inter-entity relationships are organized into *semantic subgraph* and *geometry subgraph* separately. The geometric relationships between specific nodes in *semantic sub-graph* and corresponding nodes *geometry sub-graph* are established to create linkages between *semantic subgraph* and *geometry subgraph*. The structure of SEKG is illustrated in Figure 6.

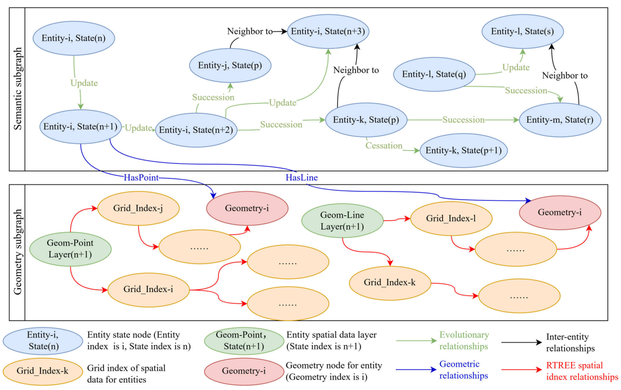

Figure 6 Framework of SEKG

As is shown in Figure 6, the *semantic subgraph* is constructed to organize and store all semantic data (entity attributes and relationships) based on a property graph. The nodes in this subgraph are Entity State Nodes representing different states of different entities. Each Entity State Node stores all attribute of an entity in corresponding time point, which is determined by its attribute of *Load time*. The relationships between different Entity State Nodes represent inter-entity relationships (e.g., *neighbor to*, *composite*, *belong to*, etc.) or evolutionary relationships (including *update*, *cessation*, *succession*) between different entity states along with temporal attributes of the relationship.

The geometry subgraph organizes spatial data layers, index grids, and entity geometries into a multi-way tree structure, implemented by property graph pointing root-to-leaves to support hierarchical management and spatial queries. The root node represents a spatial data layer, storing its name, CRS (in WKT), extent (represented by bounding box), and coordinate attribute name of geometry nodes in this layer. The internal nodes are index grids storing bounding boxes of current grid. The children of grid nodes are either subdivided grids or entity geometry nodes (leaf node of the tree) within that grid. Leaf nodes are entity geometry nodes storing geometry (in WKT) and bounding boxes. Besides linking to corresponding grid index node, each geometry node is also linked to its corresponding entity state node via a geometric relationship for coordinated state-geometry management.

Based on the integration of the two subgraphs, the SEKG is constructed. When implemented based on graph databases (e.g., Neo4j), it empowers comprehensive entity retrieval, including attributes, spatial data, evolutionary processes and relationships, by providing three fundamental search modes: node-based, property-based, and edge-based traversal.

For attribute retrieval tasks in SEKG, both geographic state nodes and relationships have their attributes. These attributes could be retrieved by simple attributed query, which is supported by all property graph databases and is usually optimized by attribute indexes (e.g., B-Tree, vector, full text search index, etc.). Furthermore, some graph databases further optimize the nodes retrieval by labelling the nodes, which divides the graph to different shards to better organize geographic nodes. With this, it could be set as the entity class name, to further improve the efficiency for node retrieval by partitioning graphs in ways that align with human cognitive intuition. Additionally, different states of entities and relationship between different states could be directly retrieved based on where clauses for attributes.

For spatial retrieval tasks, it could be accomplished by graph searching algorithms in SEKG, which connects the geometry nodes not only to state nodes with geometric relationships but also to corresponding spatial index grid nodes and spatial layer nodes with RTREE relationship. Based on this, the geometry of an entity could be retrieved by a triplet query when its state node is known or by a spatial intersection identification, which starts from the geometry layer node and finished when the node is found in the corresponding tree by comparing the bounding boxes of grid index nodes and the geometry nodes correspondingly.

For relationship retrieval tasks, the SEKG could conduct retrieval of triplets with graph searching algorithms, provided by all types of graph databases. In SEKG, entity states are connected by inter-entity relationships or evolutionary relationships. This not only provides an intuitive way to retrieve inter-entity relationships with triplets, but also capable of the evolutionary process of an entity by filtering the type of links between nodes with evolutionary relationships. Furthermore, leveraging the inherent advantages of graph-based data structures, it effortlessly enables multi-hop relationship queries. This capability provides a robust technical foundation for spatiotemporal semantic mining based on entities.

Building on this, users can effortlessly access comprehensive spatiotemporal information and dynamic evolution processes throughout an entity's full lifecycle; they can also organize versioned entity data based on temporal attributes of entities and relationships, thereby unlocking new application scenarios for entity data utilization.

4.3 Incremental update technology for SEKG

Generally, incremental updating for a knowledge graph is a crucial process for ensuring its information accurately reflects both the latest status and the full lifecycle evolution of entities and their relationships. To maintain compatibility with existing fundamental surveying and mapping data production workflows, this research extends the incremental updating techniques used for current 1:50,000 scale fundamental geographic information data. By integrating requirements outlined in the Technical Documents for New Fundamental Surveying and Mapping and Real Scene 3D China Construction, we develop an incremental updating technique specifically for geographic entity knowledge graphs. The overall technical approach is illustrated in Figure 7.

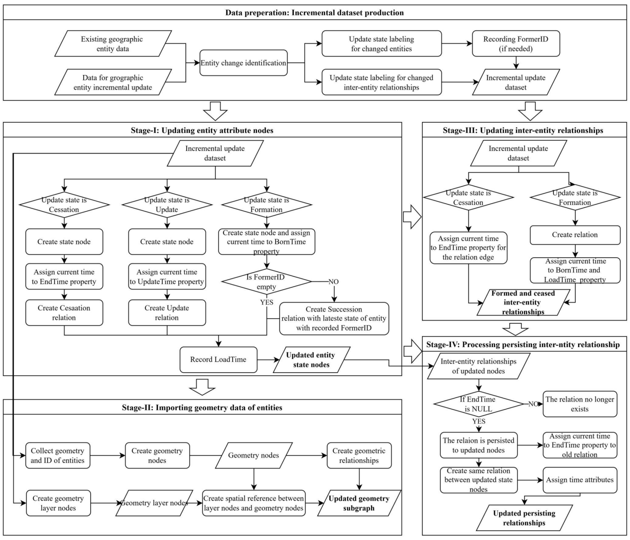

Figure 7 Incremental update technique for SEKG

The knowledge graph update process comprises a data preparation stage and four knowledge graph update stages. The data preparation is a process of data producing based on historical data and the latest production inputs. In this study, such progress was conducted based on the data product produced by projects for Fundamental Surveying and Mapping Data Production that records using incremental update flags: "Creation", "Update", and "Cessation". Simultaneously, the historical entity IDs for newly formed entities are populated according to the requirements of the "Technical Documents for New Fundamental Surveying and Mapping and Real Scene 3D China Construction". Relationship changes are primarily marked as either formation or discontinuation.

The knowledge graph update stage consists of the following four substage. (1) Incremental Update of Entity Semantic States: This stage involves two steps: First, import entities flagged as formation. If such an entity records a historical entity unique ID, it establishes a succession relationship between it and the latest state node of the corresponding entity. Subsequently, import entities with update states update, cessation, establishing the corresponding evolutionary relationship with their latest state node in the graph. (2) Incremental Update of Entity Spatial Information: this stage mainly constructs the spatial data layer node and the spatial geometry node for each entity state within the knowledge graph. For each data layer, create RTREE spatial index relation between data layer node and corresponding grid index nodes, geometry nodes based on their spatial location and extents. (3) Entity Relationship Processing: This stage mainly deals with changed relationships, including creating relationships marked as formation and assigning the end time attribute to relationships marked as discontinuation. (4) Processing persisting inter-entity relationships: This stage needs to deal with the situation that inter-entity relationships remain unchanged while corresponding entities are updated. First of all, all relationships that have not been marked as discontinuation (end time remains null) are extracted from the previous state node of updated entities. Then, create these relationships between the corresponding latest state nodes of the updated entities, along with the born time recorded by the original relationship.

5. Implementation of SEKG

This study utilizes the Fundamental Geographic Entity data from 2022, 2023, and 2024 to construct a SEKG for National Fundamental Geographic Entity using the Neo4j graph database. First of all, the data, including attribute, spatial, and relationship data, in 2022 was imported into Neo4j separately to establish the baseline SEKG of geographic entity after dealing with multiple geometric representations for a single entity through cross-table consolidation of its attributes. Subsequently, incremental data extracted from the 2023 and 2024 datasets were used to perform incremental updates on the constructed SEKG. Based on this, a multi-temporal geographic entity knowledge graph storing different states of entities and inter-entity relationships with different temporal currency was constructed. All entity information, including attribute, spatial location, extent, inter-entity relationships and evolutionary process, in any temporal state can be directly retrieved using a single Cypher query language. To better visualize the SEKG, we developed and deployed data retrieval interfaces for SEKG using Python 3.11 with packages of neo4j and fastapi to provide geographic entity data service and a visualization module for SEKG using ECharts to visualize the knowledge graph and Cesium to locate and visualize spatial information of the entities, as illustrated in Figure 8.

Figure 8 Implementation of SEKG

As is shown in Figure 8, based on the SEKG and customized data interfaces, semantic and spatial information of any state of a geographic entity could be retrieved and visualized simultaneously in the browser. When viewing the entity states with the graph, all types of existing relationship (including inter-entity relationships and evolutionary relationship if exists) of the

https://doi.org/10.5194/isprs-archives-XLVIII-4-W14-2025-445-2025 | © Author(s) 2025. CC BY 4.0 License.

selected entity state and statistics of related entity states could be retrieved directly from the SEKG, the geometry of corresponding entity state that selected was retrieved and visualized on the map simultaneously based on the bounding box and coordinates saved in corresponding geometry nodes. In this way, the SEKG provides an efficient way to visualize comprehensive information of an entity in all states through its full life cycle.

6. Discussion

Unlike conventional way that manages geographic entity data using relational database, knowledge graph adopts fundamentally distinct data organization approach with respect to conventional relational database. To evaluate whether this paradigm enhances capabilities and performance of the database, this study constructed a versioned geographic entity database using PostGIS with identical data. Entity attributes and spatial data are stored in the same spatial table for the same type of entities. The tables are hierarchically partitioned by entity classes, geometry type, and temporal validity year, while inter-entity relationships reside in separate non-spatial tables.

In terms of application capabilities, the SEKG structurally organizes and stores all entity information through semantic frameworks, offering significant advantages for semantic-oriented data storage and analysis. These benefits manifest in three key aspects:

(1) The SEKG drastic simplifies semantic query logic with graph searching algorithms. By semantically encoding entity information into property graphs, knowledge graphs dramatically streamline query syntax. For instance, retrieving a triple (subject-predicate-object) can be achieved with a single Cypher statement in Neo4j. The query session could directly return all information of related entities and their interrelationships. In contrast, SQL-based retrieval requires multiple steps with nested subqueries for multi-hop relationships as follow: querying geometric/attribute data from entity tables, joining relationship tables for relational context. In this way, the SEKG efficiently handles such traversals through graph path expansion, whereas SQL demands complex query nesting. This fundamentally accelerates semantic queries and knowledge mining and simplifies the development of application functions for users to use geographic entity data with multi-temporal and multi-dimensional information.

(2) The SEKG can model spatiotemporal evolution in the form of semantic relationships. Unlike conventional tabular-based versioned data management that stores discrete state snapshots of geographic entities and inter-entity relationships in different tables, knowledge graphs employs a multi-temporal currency hybrid storage model to uniformly organize geographic entities with varying temporal currency. Different states of entity are represented as distinct nodes with temporal attributes, the evolutionary processes are represented as edges between state nodes, which forms triplet to xx the evolutionary processes. This makes it possible to retrieve the entire evolution chain that represented by a series of triplets with a single query statement while in SQL-based retrieval demands separately retrieve data from multiple annual versions of entity tables and join the results by according the temporal information of retrieved entities, which the retrieval process becomes even more complex than that in SEKG. This further compound significant difficulties in the full lifecycle management and utilization of entities.

(3) Graph-based SEKG inherently supports semantic mining via graph algorithms. As the SEKG is constructed based on graph mode, it natively supports different graph algorithms (e.g., TransE, node2vec). Researchers can perform semantic mining analysis, such as spatiotemporal pattern mining, semantic relationship inference, dynamic evolution prediction, directly on the knowledge graph without data transformation. Conversely, tabular-based databases offer limited semantic support due to their rigid tabular structure, which struggles to support applying data mining with semantic algorithms natively. This explicitly materializes entity evolution trajectories in computable form.

In terms of performance comparison between SEKG and conventional PostGIS-based database, the SEKG based on Neo4j and conventional versioned geographic entity data based on PostGIS were constructed and deployed on same environment with 2×Intel Xeon E5-2430v2, 32GB RAM, 10TB HDD storage and Kylin Linux V10 operating system. Subsequently, five semantic retrieval scenarios and two spatial retrieval scenarios were designed for benchmark: (1) Attribute retrieval, retrieve all attributes of entity named Beijing-Fuyuan (G102) based on its entity ID. (2) Relation retrieval, relationship query between Haerbin-Tongjiang Express Way (G1011) and Mudanjiang Bridge based their entity ID. (3) Deep relation retrieval, query if Hegang-Haerbin Epxress Way (G1111) passes through Hulan River by a specific bridge based on their entity names; (4) Attribute retrieval with deep relation search, retrieve all entity attributes in scenario (3); (5) Related entity retrieval, retrieve all entities related to Yintang Trunk Canal, which is located in Tangyuan Count in Heilongjiang province, based on its entity ID; (6) Spatial retrieval (with rectangle area), query all transportation features within a 0.5-degree latitude/longitude bounding box centered on Suihua City (approximately 5,000 km²). (7) Spatial retrieval (with complex polygon area), the query region is counstructed by merging the polygons of Wucui District and Youhao District in Yichun City into a single search region and adjusting its spatial location and extent through scaling, moving and rotating to achieve approximate same query results with query area in (6).

With all the samples prepared for benchmarking, all queries were executed via official Python drivers on a workstation connected to the server that the Neo4j and PostGIS databases are distributed on. To eliminate cold-start interference, warm-up queries (using disjoint data) preceded 30 times test runs with pre-query cache cleared for each time. Average query times of different scenarios are documented in Table 1.

Retrieval scenarios	Number of records retrieved	Time cost (seconds)	
		Neo4j	PostGIS
Attribute retrieval	1	0.0406	0.0327
Relation retrieval	1	0.0491	0.0625
Deep relation retrieval	2	0.0768	0.1593
Attribute retrieval with deep relation search	6	0.0823	0.7604
Related entity retrieval	12	0.0566	1.0958
Spatial retrieval (with rectangle area)	3462	5.8807	0.1120
Spatial retrieval (with complex polygon area)	3462	10.1559	0.1543

Table 1 Retrieval time comparison between Neo4J and PostGIS

As shown in Table 1, Neo4j and PostGIS exhibit comparable performance in basic queries, such as retrieving a single entity attribute or relationship via attribute filtering, with both returning results within approximately 50 milliseconds. However, their performance diverges significantly in complex semantic queries and spatial queries. For complex semantic scenarios (e.g., deep relationship traversal, querying attributes of deeply associated entities, or retrieving attributes across related entities), Neo4j consistently outperforms PostGIS. This advantage arises because such queries require accessing multiple relationship records and attributes from diverse entity types. Neo4j efficiently retrieves this information via direct pointer or linked list traversal, whereas PostGIS must first query relationship paths from association tables using entity IDs, then perform conditional searches across multiple entity tables to gather attributes. This process incurs substantial computational loads, particularly in the case of *Related entity retrieval*, where associated entities reside across seven distinct tables, resulting in significantly higher query times compared to other semantic queries. Conversely, PostGIS demonstrates superior performance for spatial queries, significantly outpacing Neo4j. This performance gap widens as the complexity of the query geometry increases. The primary reason lies in Neo4j's implementation: its R-TREE spatial index still relies on graph traversal for spatial lookups. When datasets cover large spatial extents, the computational cost associated with traversal escalates markedly. Consequently, Neo4j lags behind PostGIS, which leverages its highly optimized GiST index specifically designed for efficient spatial retrieval. In summary, the SEKG constructed using Neo4j facilitates efficient semantic search operations through graph algorithm implementations while necessitating computationally expensive in spatial approximations, consequently undermining efficiency in spatial query through graph-based path-finding mechanisms.

7. Conclusion

This study, grounded in the analysis on spatiotemporal evolutionary scenarios of geographic entities and inter-entity relationships, proposes a Spatiotemporal Evolution Knowledge Graph along with incremental updating techniques for geographic entities. This provides a unified framework for organizing, storing, and managing comprehensive information throughout the entire life cycle of geographic entities. The capability and performance of the SEKG was then validated by comparison with the conventional relational database methodologies. The validation shows that the SEKG approach achieves the integrated organization and storage of multi-dimensional geographic entity information (including attributes, spatial locations, spatial extents, inter-entity relationships, and evolutionary processes). It supports the coordinated retrieval, representation, and dynamic updating of heterogeneous spatial and semantic information, providing technical support for full lifecycle management of geographic entities and entity-based spatiotemporal modeling and dynamic simulation. Furthermore, the knowledge graph serves as the foundational framework of this research, offering a graph-native infrastructure naturally compatible with efficient graph analysis, semantic computation, and knowledge mining algorithms. This core advantage establishes the technical foundation for entity-based spatiotemporal knowledge discovery research, such as spatiotemporal correlation analysis, evolution mechanism investigation, and hidden knowledge extraction.

Although the SEKG proposed in this study effectively manages geographic entities throughout their life cycle and enables collaborative storage of spatiotemporal information, its spatial data support remains primarily limited to 2D data, with performance in spatial operations still lagging behind traditional relational spatial databases; future research should therefore focus on designing a unified 2D/3D storage model for knowledge graph construction, exploring efficient graph-based spatial indexing methods, and constructing a hybrid database that integrates 2D/3D geometries with semantics to better support entity-based data management, spatiotemporal modelling and knowledge discovery.

References

Abu-Salih, B., 2021: Domain-specific knowledge graphs: A survey. *Journal of Network and Computer Applications*, *185*, 103076. doi.org/10.1016/j.jnca.2021.103076

Arvor, D., Durieux, L., Andrés, S., and Laporte, M.-A., 2013: Advances in Geographic Object-Based Image Analysis with ontologies: A review of main contributions and limitations from a remote sensing perspective. *ISPRS Journal of Photogrammetry and Remote Sensing*, *82*, 125–137. doi.org/10.1016/j.isprsjprs.2013.05.003

Bateman, J., and Farrar, S. (2004). Towards a generic foundation for spatial ontology. Proceedings of the 3rd International Conference on Formal Ontology in Information Systems (FOIS-04).

Breunig, M., Bradley, P. E., Jahn, M., Kuper, P., Mazroob, N., Rösch, N., Al-Doori, M., Stefanakis, E., and Jadidi, M., 2020: Geospatial Data Management Research: Progress and Future Directions. *Isprs International Journal of Geo-Information*, *9*(2). 10.3390/ijgi9020095

Ji, S., Pan, S., Cambria, E., Marttinen, P., and Yu, P. S., 2022: A Survey on Knowledge Graphs: Representation, Acquisition, and Applications. *IEEE Transactions on Neural Networks and Learning Systems*, *33*(2), 494–514. 10.1109/TNNLS.2021.3070843

Lawley, C. J. M., Gadd, M. G., Parsa, M., Lederer, G. W., Graham, G. E., and Ford, A., 2023: Applications of Natural Language Processing to Geoscience Text Data and Prospectivity Modeling. *Natural Resources Research*, *32*(4), 1503–1527. 10.1007/s11053-023-10216-1

Ma, X., 2022: Knowledge graph construction and application in geosciences: A review. *Computers & Geosciences*, *161*. 10.1016/j.cageo.2022.105082

Mai, G., Janowicz, K., Cai, L., Zhu, R., Regalia, B., Yan, B., Shi, M., and Lao, N., 2020: SE‐KGE: A location‐aware Knowledge Graph Embedding model for Geographic Question Answering and Spatial Semantic Lifting. *Transactions in GIS*, *24*(3), 623–655. 10.1111/tgis.12629

Shen, Y., 2018: A New Process-Oriented and Spatiotemporal Data Model for Gis Data. *The International Archives of the Photogrammetry, Remote Sensing and Spatial Information Sciences*, *XLII-3*, 1539–1541. 10.5194/isprs-archives-XLII-3-1539-2018

Shi, L., Lan, X., Xiao, M., and Liu, N., 2024: Geographical Entity Management Model Based on Multi-Classification. *Isprs International Journal of Geo-Information*, *13*(9). 10.3390/ijgi13090334

Wang, S., Zhang, X., Ye, P., Du, M., Lu, Y., and Xue, H., 2019: Geographic Knowledge Graph (GeoKG): A Formalized Geographic Knowledge Representation. *Isprs International Journal of Geo-Information*, *8*(4), 184. 10.3390/ijgi8040184

Xiao, J., Li, H., Li, P., Wang, S., and Chen, K., 2021: Connotation,Classification and Grading of 3D Real Scene. *Urban Geotechnical Investigation & Surveying*(5), 5–10.

Xue, C., Wu, C., Liu, J., and Su, F., 2019: A Novel Process-Oriented Graph Storage for Dynamic Geographic Phenomena. *Isprs International Journal of Geo-Information*, *8*(2). 10.3390/ijgi8020100

Zhang, L., Ding, Y., Zhu, Q., Chen, X., Li, H., and Wang, w., 2024: A Holistic Multi-domain Association Model for Railway Geographic and Geologic Data. *Geomatics and Information Science of Wuhan University*, *49*(6), 1018–1027. 10.13203/j.whugis20220010

Zhang, Y., and Tian, H., 2025: Engineering modeling technology of fundamental geo-entity. *Bulletin of surveying and mapping 0*(05), 159–164. 10.13474/j.cnki.11-2246.2025.0526

Zheng, K., Xie, M. H., Zhang, J. B., Xie, J., and Xia, S. H., 2021: A knowledge representation model based on the geographic spatiotemporal process. *International Journal of Geographical Information Science*, *36*(4), 674–691. 10.1080/13658816.2021.1962527

Zheng, Y., Tian, H., Zhang, y., and Li, X., 2023: Design and practice on national basic grographic entity construction technology. *Science of Surveying and Mapping 48*(03), 25–29. 10.16251/j.cnki.1009-2307.2023.03.004

Zhu, J., Fu, Y., Zhou, J., and Chen, D., 2025: A Temporal Knowledge Graph Generation Dataset Supervised Distantly by Large Language Models. *Scientific Data*, *12*(1), 734. 10.1038/s41597-025-05062-0