Temporal Data Containers: A Semantic Framework for Longitudinal Spatial Data Integration and Quality Assessment

Andrea Roberta Costagliola, Marco Montanari, Paolo Bellavista
Department of Computer Science and Engineering, University of Bologna
Alma Mater Studiorum - University of Bologna
Bologna, Italy
0009-0009-0119-2597, 0000-0001-5026-6083, 0000-0003-0992-7948

1. Introduction and Motivation

One of the most critical yet under-addressed challenges in the development of Urban Digital Twins (UDTs) is the semantic management of spatial-temporal continuity in heterogeneous datasets. In domains such as urban history, planning, and governance, the ability to reconstruct how places evolve across time—through processes like territorial splits, mergers, renaming, or reclassification—is essential to ensure data consistency and interpretability (1; 5).

However, most current urban data integration approaches fall short in this respect. While systems such as GIVA (1) and Data Polygamy (4) enable spatial-temporal analysis through visualization and statistical overlays, they lack explicit mechanisms to encode and reason over temporal transformations. Similarly, semantic frameworks such as Urban Knowledge Graphs (6) and the Geographic Evolutionary Knowledge Graph (GEKG) (7) provide rich vocabularies but rarely model entity-level lineage or formal transformation rules. Approaches like Theseus (8) automate change detection across datasets but are not integrated into collaborative platforms that support temporal reasoning. This leaves a critical gap in the ability to support reliable longitudinal reconstructions and transformation-aware integration.

To address these limitations, we build upon our previously introduced extension of the GlassBox model, initially designed for modular and observable urban data processing. In this work, we enhance that model with the notion of Temporal Data Containers (TDCs)—semantic abstractions that enable the incremental integration and reconstruction of spatial entities across time. TDCs encode formal transformation rules (e.g., splits, merges, renamings), provenance metadata, and data quality descriptors along three dimensions: spatial precision, temporal granularity, and semantic completeness. The framework is implemented through a Semantic MediaWiki infrastructure, extended with inferential modules and dynamic query capabilities. This environment supports the collaborative construction of time-aware urban knowledge graphs that bridge historical and contemporary data, ensuring both interoperability and interpretability.

To validate our proposal, we apply it to the longitudinal reconstruction of administrative entities in Italy using datasets published by the National Institute of Statistics (ISTAT).

Identify applicable funding agency here. If none, delete this.

I. 2. Concept and Model

Temporal Data Containers act as semantic units that encapsulate the geometry, validity period, provenance, and transformation lineage of spatial entities across time. Each TDC represents a temporally bounded instance of an urban entity and is formally linked to its predecessors and successors through explicitly modeled transformation rules.

This extended framework enables time-aware integration of spatial data by supporting semantic alignment across multiple dataset versions. In particular, it allows entities describing the same spatial feature (e.g., a municipality) to be semantically connected even in the presence of:

- heterogeneous formats and evolving data standards,
- misaligned geometries or inconsistent identifiers,
- partial, ambiguous, or missing attribute sets.

A representative use case involves administrative boundary datasets, such as those periodically released by national statistical institutes. These datasets often reflect complex changes over time—including splits, mergers, renamings, and geometric refinements—that are semantically captured through transformation rules (e.g., `SplitRule`, `MergeRule`, `RenameRule`).

Each dataset version is captured as a node in a semantically indexed knowledge graph, annotated with temporal metadata, spatial relations (e.g., overlaps, containment, adjacency), and provenance descriptors.

By integrating transformation semantics and temporal reasoning into the core data model, the system enhances interpretability and enables quality-aware analyses, including change detection, historical validation, and coverage assessment. These capabilities form the foundation for building time-aware Digital Twins and conducting longitudinal studies in urban and territorial governance.

Each Temporal Data Container also stores quality descriptors that contribute to the notion of *information density*, defined along three key dimensions: (i) **spatial precision**, indicating the geometric fidelity of the representation; (ii) **temporal granularity**, measuring the frequency and consistency of updates; and (iii) **semantic completeness**, reflecting the richness and coherence of descriptive attributes. These dimensions inform validation and support confidence-weighted reconstructions.

3. Use Case: Tracing the Evolution of Italian Municipalities

To validate our framework, we developed a prototype focused on the longitudinal reconstruction of Italian municipalities using datasets from the ISTAT, complemented by historical shapefiles and scanned maps. The goal is to track how spatial units—such as municipalities—have transformed across decades through administrative changes like splits, mergers, renamings, and boundary adjustments.

The system ingests each dataset release and semantically aligns it using a Semantic MediaWiki infrastructure extended with custom templates, rule-based transformation modules, and an urban ontology. Temporal relationships between entities are encoded through rules such as `SplitRule` or `MergeRule`, enabling users to issue historical queries dynamically.

For example, the platform can answer complex questions such as: ”What was the municipal structure of Emilia-Romagna in 1901?”

These capabilities support digital heritage studies, historical validation, and time-aware integration with socio-economic data layers.

4. Quality Assessment and Information Density

As introduced in Section 2, each Temporal Data Container (TDC) encapsulates quality descriptors that contribute to the concept of *information density*. This metric supports the evaluation of how well a given spatial entity is documented across time, based on its geometric accuracy, update frequency, and semantic richness.

Rather than relying on a single dataset release, the system aggregates all temporally linked instances to assess cumulative quality. For example, while a location like the Colosseum may be well-known today, earlier data may include misaligned geometries or missing attributes. The system can thus identify what is reliably known and where uncertainty or gaps persist.

These quality signals inform rule-based reconstructions, guide confidence-weighted historical queries, and support targeted data improvement across time periods or regions.

5. Conclusive Remarks and Future Work

Our framework builds on the extended GlassBox model to provide a unified semantic approach for spatio-temporal analysis across both legacy and contemporary datasets. It enables longitudinal studies without requiring rigid schema harmonization, promotes the reuse of heterogeneous public datasets through semantic linking, and supports analytical workflows in domains such as urban history, planning, governance, and digital heritage.

Although our prototype focuses on a national case study, the framework is explicitly designed to generalize across spatio-temporal domains. Its semantic architecture—based on Temporal Data Containers, transformation rules, and rule-based reasoning—can be extended to multiple application areas such as environmental monitoring, cadastral systems, infrastructure management, and smart mobility.

Future work will explore the integration of additional data sources and the application of the framework to new domains, including smart city services and linked environmental datasets. We also plan to enhance interoperability with established standards such as GeoSPARQL and Linked Open Data vocabularies, ensuring compatibility with semantic web infrastructures. Additionally, we aim to incorporate spatial and temporal uncertainty modeling into the reasoning layer, enabling confidence-weighted reconstructions and improving robustness in the presence of ambiguous or incomplete records.

References

[1] I. F. Cruz, V. R. Ganesh, C. Caletti, and P. Reddy, “GIVA: a semantic framework for geospatial and temporal data integration, visualization, and analytics,” in *Proc. 21st ACM SIGSPATIAL Int. Conf. Advances in Geographic Information Systems (SIGSPATIAL’13)*, Orlando, FL, USA, 2013, pp. 544–547.

[2] Y. Zhu and J. Ferreira, “Data integration to create large-scale spatially detailed synthetic populations,” in *Planning Support Systems and Smart Cities*, S. Geertman, J. Ferreira Jr., R. Goodspeed, and J. Stillwell, Eds. Cham, Switzerland: Springer, 2015, pp. 121–141.

[3] A. Psyllidis, “Ontology-based data integration from heterogeneous urban systems: A knowledge representation framework for smart cities,” in *Proc.*, 2015. [Online].

[4] F. Chirigati, H. Doraiswamy, T. Damoulas, and J. Freire, “Data Polygamy: the many-many relationships among urban spatio-temporal data sets,” in *Proc. 2016 Int. Conf. Management of Data (SIGMOD ’16)*, San Francisco, CA, USA, 2016, pp. 1011–1025.

[5] W. Kuhn, “Core concepts of spatial information for transdisciplinary research,” *Int. J. Geogr. Inf. Sci.*, vol. 26, pp. 2267–2276, Dec. 2012.

[6] Y. Liu, J. Ding, Y. Fu, and Y. Li, “UrbanKG: An Urban Knowledge Graph System,” *ACM Trans. Intell. Syst. Technol.*, vol. 14, no. 4, article 60, pp. 1–25, Aug. 2023. doi:10.1145/3588577.

[7] Zheng, Kun; Xie, Ming Hui; Zhang, Jin Biao; Xie, Juan; Xia, Shu Hao. *A Knowledge Representation Model Based on the Geographic Spatiotemporal Process*. International Journal of Geographical Information Science, vol. 36, pp. 674–691, 2021.

[8] C. Bernard, M. Viry *et al.*, “Theseus: Tracking spatio-temporal administrative changes using knowledge graphs,” in *Proc. of Semantic Big Data Workshop*, 2022