From Sensor Streams to Health Narratives: Semantic Data Storytelling Using Wearable Knowledge Graphs

Asara Senaratne*, Leelanga Seneviratne

College of Science and Engineering, Flinders University, Adelaide, South Australia.

Abstract

Wearable devices continuously collect rich streams of physiological, behavioural, and geospatial data, yet these observations are typically presented as isolated measurements or simple dashboards that provide limited support for understanding the temporal context of an individual's health. This paper presents a semantic framework for data-driven health storytelling that transforms heterogeneous wearable-device data into coherent, interpretable health narratives. The proposed pipeline integrates Apple Health records and GPS trajectories into a knowledge graph based on the Chronic Observation and Progression Events (COPE) ontology, semantically linking activities, physiological observations, locations, and temporal relationships. Rather than treating wearable measurements as disconnected data points, the framework organises them into meaningful health trajectories that capture how physiological responses evolve throughout daily activities. Using a real-world case study, we demonstrate how semantic integration enables the construction of contextual narratives by combining heart rate, movement, energy expenditure, cadence, and spatial information into a unified representation. Interactive visualisations illustrate how semantically connected observations reveal transitions between activity phases, highlight contextual factors influencing physiological changes, and support intuitive exploration of personal health data. Compared with conventional dashboards that display individual metrics independently, the proposed approach provides a richer, explainable view of health events by preserving temporal and contextual relationships across multiple data modalities. This work demonstrates the potential of semantic technologies to support data-driven storytelling in personal health, providing an interpretable foundation for longitudinal health exploration, personalised wellbeing applications, and future integration with clinical knowledge graphs and explainable health analytics.

Keywords

Ontology, Health Trajectory, Wearable Device Data, Temporal Events

CEUR-WS.org/Vol-4313/dds-paper3.pdf

1. Introduction

Wearable devices such as smartwatches and fitness trackers continuously collect large volumes of physiological, behavioural, and geospatial data, providing unprecedented opportunities for monitoring personal health and wellbeing [1]. Modern wearable platforms capture diverse observations including heart rate, physical activity, sleep, energy expenditure, mobility, and location. Although these data are routinely presented through dashboards and summary statistics, they are often visualised as isolated measurements, making it difficult to understand how multiple observations interact over time or to explain the progression of an individual's health in a meaningful context. Furthermore, existing devices typically perform anomaly detection [2, 3, 4] on individual vital-sign streams in isolation, without considering the interrelationships and interactions among multiple vital signals. Consequently, users and healthcare practitioners must manually interpret fragmented information to identify behavioural patterns, understand physiological responses, or assess lifestyle changes [5].

Data-driven storytelling has emerged as an effective paradigm for transforming complex datasets into coherent narratives that improve interpretation, communication, and decision making. Rather than presenting disconnected visualisations, storytelling combines multiple sources of evidence into contextual explanations that describe what happened, when it occurred, and how different observations relate to one another [6]. For wearable health data, this requires integrating heterogeneous measurements

Data-Driven Storytelling: Bridging Knowledge Graphs, GenAI, and Narrative (DDS2026), in ISWC 2026 Workshops Joint Proceedings, October 25–26, 2026, Bari, Italy

*Corresponding author.

asara.senaratne@flinders.edu.au (A. Senaratne); leelanga.seneviratne@flinders.edu.au (L. Seneviratne)

0000-0002-3080-7847 (A. Senaratne); 0000-0002-5761-3014 (L. Seneviratne)

© 2026 Copyright for this paper by its authors. Use permitted under Creative Commons License Attribution 4.0 International (CC BY 4.0).

CEUR Workshop Proceedings ceur-ws.org ISSN 1613-0073

published 2026-11-05

across temporal, physiological, behavioural, and spatial dimensions while preserving the relationships that give rise to meaningful health events.

Semantic technologies provide a natural foundation for constructing such narratives [7, 8, 9]. Knowledge graphs explicitly represent entities, events, temporal relationships, and contextual information, enabling heterogeneous observations to be linked into an interpretable representation of an individual's evolving health. By modelling observations semantically, physiological measurements can be connected with activities, locations, trajectories, and temporal progression, allowing isolated sensor readings to be interpreted as components of a broader health story rather than independent data points.

Therefore, in this research paper, we investigate the research question; *how can knowledge graphs support explainable, data-driven storytelling from heterogeneous wearable device data?* To address this question, we present a semantic framework that transforms wearable-device exports and associated geospatial trajectories into a knowledge graph using the Chronic Observation and Progression Events (COPE) ontology[1] [1]. The resulting representation integrates physiological observations, activities, spatial trajectories, and temporal relationships into a unified semantic model that supports the construction of contextual health narratives. Through a real-world case study using Apple Health data, we demonstrate how semantically linked observations can be visualised as coherent health trajectories, revealing relationships between physiological responses, movement patterns, and environmental context that are difficult to interpret using conventional dashboards alone.

Hence, the primary **contribution** of this work is; a *semantic modelling pipeline* that links physiological measurements, activities, and geospatial trajectories using the COPE ontology to preserve temporal and contextual relationships across multiple data modalities.

By positioning semantic modelling as an enabler of explainable health narratives rather than simply a data transformation mechanism, through this work we aim to demonstrate how knowledge graphs can support more interpretable exploration of personal health data and provide a foundation for future personalised health analytics, digital twins in healthcare, clinical integration, and explainable decision-support systems.

2. Literature Review

Semantic interoperability represents a fundamental requirement for healthcare digital twins (HDTs), as they require continuous integration of heterogeneous data sources including wearable sensing devices, electronic health records (EHRs), biomedical repositories, and patient-generated health data. Unlike conventional data integration approaches, semantic technologies provide machine-interpretable representations that enable contextual understanding, reasoning, and knowledge-driven decision support. Therefore, healthcare digital twins increasingly rely on ontologies, semantic information models, and knowledge graphs to establish a common representation layer across diverse healthcare ecosystems [10, 11]. This section reviews the major semantic standards and knowledge representation approaches that support HDT development. Specifically, we discuss semantic sensor network ontologies for representing physical and IoT environments, healthcare interoperability standards for clinical information exchange, and knowledge graph based approaches for personalised health modelling and patient digital twins.

2.1. Semantic Sensor Networks and IoT Ontologies

The semantic representation of sensing devices, observations, and actuation processes forms the foundation for real-time monitoring in healthcare digital twins. Wearable devices, ambient sensors, and medical IoT platforms continuously generate heterogeneous observations that require semantic descriptions to enable interoperability, discovery, and reasoning across distributed healthcare environments.

The W3C Semantic Sensor Network (SSN) ontology provides a comprehensive semantic model for describing sensors, observations, sampling processes, and actuation capabilities [12]. To support

[1] https://hzent.github.io/COPE/CopeDashboard.html

lightweight adoption and improve modularity, the Spatial Data on the Web Working Group introduced the Sensor, Observation, Sample, and Actuator (SOSA) ontology as a lightweight core aligned with SSN [13]. SOSA establishes the fundamental semantic relationship between sensing entities and observed phenomena: `sosa:Sensor` ⟶`sosa:Observation` ⟶`sosa:ObservableProperty`, where an observation represents a measurement generated by a sensor over an observable property associated with a specific `sosa:FeatureOfInterest`. Similarly, physical interventions and control operations are represented through the actuation model: `sosa:Actuator` ⟶`sosa:Actuation` ⟶`sosa:ActuatableProperty`; The SSN extension enriches this lightweight representation by introducing additional descriptions related to sensing capabilities, deployment platforms, operating conditions, and measurement characteristics through alignment with foundational ontologies such as DOLCE. Within healthcare digital twins, SSN/SOSA enables semantic integration of physiological observations, environmental context, and device metadata generated from heterogeneous sensing infrastructures.

For resource-constrained IoT environments, lightweight semantic models have been proposed to reduce computational overhead while maintaining interoperability. The IoT-Lite ontology provides a simplified extension of SSN by focusing on essential IoT concepts, including devices, services, measurements, and spatial coverage [14]. This lightweight representation is particularly relevant for edge-based healthcare applications where semantic processing must be performed under limited computational and energy constraints.

Similarly, the Smart Applications REFerence Ontology (SAREF), developed within the ETSI standardisation framework, provides an industry-oriented semantic model for representing smart devices and their interactions [15]. SAREF defines concepts including;1 `saref:Device`, `saref:Function`, `saref:State`, measurements, units, and commands. Healthcare-specific extensions, such as SAREF4EHA, extend this semantic model towards e-health and ageing well applications by supporting representation of healthcare devices, assisted living environments, and medical observations [16].

The Open Geospatial Consortium (OGC) Sensor Model Language (SensorML) provides a standardised representation of sensor systems, processing workflows, and measurement procedures [17]. Although originally designed for geospatial sensor infrastructures, SensorML has been widely adopted within semantic IoT research by mapping sensor descriptions into RDF-based models such as SSN/SOSA. These mappings enable semantic sensor discovery, automated composition of sensing services, and interoperable Sensor Web Enablement (SWE) architectures.

2.2. Distinction from Conventional Fitness Visualisations

Commercial fitness platforms already provide effective visualisations that combine measurements such as heart rate, cadence, speed, and GPS trajectories [18, 19]. The contribution of the proposed approach is therefore not the creation of another integrated workout dashboard. Instead, it concerns the semantic representation underlying such views. Commercial applications generally expose users to application-specific representations and predefined visualisations [20], whereas the proposed framework represents observations, activities, trajectories, temporal relationships, and contextual properties explicitly in an interoperable knowledge graph. This representation allows the same underlying data to be queried across modalities, enriched with external semantic vocabularies, and reused to construct alternative narratives or analytical views [21]. For example, a query can identify periods of elevated cardiovascular response and subsequently retrieve the associated activity, location, trajectory, and neighbouring observations. Thus, the novelty lies in the semantic integration and queryable representation that supports explainable storytelling, rather than in reproducing functionality already available in fitness applications.

2.3. Healthcare Semantic Standards and Biomedical Ontologies

Healthcare digital twins require semantic models capable of representing both clinical knowledge and continuously generated physiological observations. Existing healthcare standards provide complementary capabilities: terminology systems define clinical meaning, while information models specify how

healthcare data are structured and exchanged.

HL7 Fast Healthcare Interoperability Resources (FHIR) has become a widely adopted standard for exchanging healthcare information by defining structured representations of clinical entities including patients, observations, conditions, and medication information [22]. Although FHIR resources are commonly exchanged using JSON or XML representations, the FHIR RDF specification transforms these resources into semantic graphs using RDF and OWL-based descriptions [23]. This transformation enables graph-based querying, semantic reasoning, and integration with external knowledge sources while maintaining compatibility with existing healthcare information systems.

SNOMED CT provides a comprehensive clinical terminology based on description logic, enabling formal representation of diseases, anatomical structures, procedures, and clinical findings through hierarchical and semantic relationships [24]. Its logical structure supports reasoning over clinical concepts and facilitates integration of heterogeneous healthcare datasets. LOINC provides standard identifiers for laboratory tests, physiological measurements, and clinical observations, enabling consistent interpretation of measurement data across healthcare systems [25]. In contrast, the International Classification of Diseases (ICD) [2], maintained by the World Health Organization, provides disease classification schemes primarily supporting epidemiological analysis, reporting, and healthcare administration.

The Observational Health Data Sciences and Informatics (OHDSI) initiative introduced the Observational Medical Outcomes Partnership Common Data Model (OMOP-CDM) to standardise heterogeneous healthcare databases into a common representation for large-scale observational research [5]. The model provides structured representations of clinical events, including; `PERSON`, `DRUG_EXPOSURE`, `CONDITION_OCCURRENCE`. Complementary to healthcare information models, biomedical ontologies developed through the Open Biomedical Ontologies (OBO) Foundry provide formal representations of biological and clinical knowledge [26]. Examples include the Human Phenotype Ontology (HPO), Gene Ontology (GO), and Disease Ontology (DO), which support phenotype matching, biological reasoning, and genomic knowledge integration.

2.4. Knowledge Graphs for Personal Health

Knowledge graphs (KGs) provide a semantic integration framework for representing heterogeneous healthcare information as interconnected entities and relationships. Within healthcare digital twins, KGs act as a knowledge integration layer that connects sensor observations, clinical records, biomedical knowledge, and patient-specific context.

Personal Health Knowledge Graphs (PHKGs) integrate patient-generated health data, wearable sensor observations, lifestyle information, and clinical records into a unified semantic representation. By modelling healthcare information as RDF graphs, PHKGs enable contextual reasoning over physiological states, behavioural patterns, environmental conditions, and treatment outcomes [6]. This capability is particularly important for HDTs, where heterogeneous observations must be continuously incorporated to maintain an up-to-date representation of an individual's health status.

A Patient Digital Twin (PDT) represents a continuously updated computational model of an individual's physiological, behavioural, and clinical state. Semantic knowledge graphs provide the interoperability layer required to integrate heterogeneous sources including IoT telemetry, EHR data, biomedical models, and intervention histories. Through continuous semantic integration and reasoning, PDTs can support disease progression modelling, adverse event prediction, treatment pathway evaluation, personalised intervention planning.

Biomedical knowledge graphs integrate diverse biological and clinical knowledge sources, including molecular interactions, drug-target relationships, clinical evidence, and longitudinal patient trajectories. These representations provide a foundation for precision medicine by enabling reasoning across multiple biological and clinical dimensions [27]. In precision healthcare applications, biomedical knowledge graphs support patient stratification, disease subtype identification, personalised treatment recommendation, temporal analysis of genomic and phenotypic information.

[2] https://icd.who.int/en/

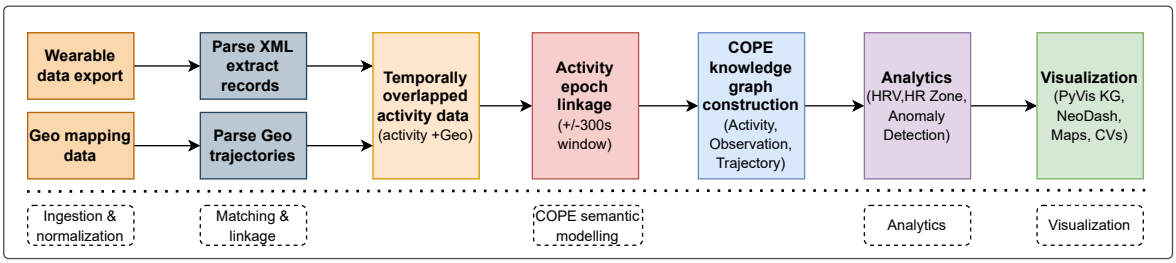


Figure 1: Pipeline for semantically modelling wearable-device data using COPE ontology.

Overall, semantic standards and knowledge representation technologies provide the foundation for interoperable [28] and intelligent healthcare digital twins. By combining sensor ontologies, healthcare information models, biomedical vocabularies, and knowledge graphs, HDTs can achieve continuous data integration, contextual reasoning, and personalised decision support across complex healthcare ecosystems [29].

However, to the best of our knowledge, only a limited number of existing ontologies attempt to semantically annotate lifestyle or activity related data in a temporal manner. For instance, eCoaching provides semantic descriptions of activity levels, while the Dietary Lifestyle Ontology (DILON) focuses on questionnaire-based dietary information [30]. HeLiS and its extension FuS-KG [31] model food and physical activity data. A key limitation across these ontologies, however, is their inability to represent the progressive and evolving nature of personal health behaviours over time [7]. In the biomedical domain, several well established ontologies capture foundational biological and clinical semantics, for example LOINC [32], the Gene Ontology, and SNOMED CT [33]. While these ontologies provide extensive terminology for clinical observations, laboratory tests, and biological attributes, they do not encode the temporal progression of events in a way that supports trajectory-based reasoning. The Chronic Observation and Progression Events (COPE) [1] ontology (namespace: https://purl.archive.org/cope#) addresses these limitations by introducing a temporally rich, progression oriented semantic framework specifically designed to represent personal health data. By linking personal attributes, temporal progression, and observation-level semantics, COPE provides a comprehensive framework for representing and reasoning about an individual's evolving health profile. Therefore, we use COPE in this study to semantically model wearable device data.

3. Semantic Framework for Data-Driven Health Storytelling

This research proposes a semantic modelling pipeline that transforms heterogeneous wearable-device data into contextual health narratives through semantic integration, temporal modelling, and interactive visualisation as shown in Figure 1. Rather than treating wearable measurements as isolated sensor readings, the framework organises physiological observations, physical activities, and geospatial trajectories into a unified knowledge graph that captures how an individual's health evolves throughout daily activities.

Let $\mathscr{D} = \{D_w, D_g\}$ denote the heterogeneous wearable dataset, where D_w represents wearable observations (e.g., exported from Apple Health) and D_g represents GPX trajectory data.

- Each wearable observation is formally defined as a tuple $o_i = (t_i, p_i, v_i, u_i)$, where t_i is the timestamp, p_i is the observed property (heart rate, cadence, etc.), v_i is the recorded value, and u_i is its measurement unit.
- Each GPS trajectory point is represented as $g_j = (t_j, \phi_j, \lambda_j, h_j)$, where t_j is the timestamp, ϕ_j denotes latitude, λ_j longitude, and h_j elevation.

The primary objective of the framework is to construct a semantic knowledge graph $G = (V, E)$ that represents contextual health narratives by integrating physiological observations, activities, and spatial

trajectories. The complete transformation pipeline is expressed as a composite mapping $f : \mathscr{D} \to G$: $f = f_6 \circ f_5 \circ f_4 \circ f_3 \circ f_2 \circ f_1$ where each composite function f_k corresponds to a sequential stage of the framework: (1) data acquisition, (2) data normalisation, (3) temporal alignment, (4) knowledge graph construction, (5) semantic enrichment, and (6) narrative generation and visualisation.

Unlike conventional wearable analytics pipelines, this framework explicitly models relationships among observations across temporal, physiological, behavioural, and spatial dimensions. These semantic relationships form the basis for constructing explainable health narratives that describe not only *what* occurred, but also *when*, *where*, and under which contextual conditions physiological changes were observed.

3.1. Data Acquisition (f_1)

The framework begins by acquiring heterogeneous data generated by wearable ecosystems. In this implementation, Apple Health exports serve as the primary data source, although the architecture is intentionally designed to remain independent of any specific wearable platform. Two categories of information are extracted during stage $f_1 : \mathscr{D} \to (O, T)$:

1. Physiological and behavioural observations exported as XML records, parsed into structured observation records $O = \{o_1, o_2, \dots, o_n\}$. These contain measurements from multiple sensors, including heart rate, heart rate variability, walking speed, cadence, energy expenditure, distance travelled, stand time, sleep duration, respiratory rate, oxygen saturation, and activity summaries.
2. Geospatial trajectories recorded as GPX files, parsed into trajectory sets $T = \{g_1, g_2, \dots, g_m\}$. These provide timestamped latitude and longitude coordinates describing the spatial path of individual workouts.

Because these datasets are generated independently, they exhibit heterogeneous schemas, sampling frequencies, timestamp resolutions, and measurement units. Unlike relational representations, observations originating from different sensing modalities remain independent at this initial stage.

3.2. Data Normalisation (f_2)

Following ingestion, all observations undergo a normalisation process f_2 to establish a consistent temporal and semantic representation.

Timestamps originating from different devices or sampling frequencies are converted into a unified temporal domain $\tau : \mathbb{T} \to \mathbb{T}_{\text{ISO}}$, such that $t_i' = \tau(t_i)$. Each observation's timestamp is updated accordingly: $o_i = (t_i', p_i, v_i, u_i)$ Measurement units are standardised using a mapping function $\eta : \mathscr{U} \to \mathscr{U}_s$, such that $u_i' = \eta(u_i)$ converts device-specific units into standardized units. Duplicated records and incomplete measurements are removed during this preprocessing step, resulting in the normalised dataset $O' = \{o_1', o_2', \dots, o_n'\}$.

3.3. Temporal Alignment (f_3)

The framework performs temporal alignment between normalised observations O' and activity sessions $A = \{a_1, a_2, \dots, a_k\}$ extracted from the wearable platform. Each activity session a_j defines an activity interval bounded by a start and end time: $a_j = (t_j^{\text{start}}, t_j^{\text{end}})$ An observation o_i is associated with an activity a_j whenever its timestamp t_i falls within the activity interval or a configurable temporal tolerance window δ (set to 300 s in this study): $t_j^{\text{start}} - \delta \le t_i \le t_j^{\text{end}} + \delta$ This defines an assignment function $\alpha(o_i) = a_j$. Similarly, GPS trajectory points g_i are aligned to activities via $\beta(g_i) = a_j$.

Spatial trajectories extracted from GPX files are thus aligned with activity intervals, allowing each physiological observation to inherit contextual information describing *where* it occurred during a workout. This temporal alignment establishes the chronological foundation required for constructing coherent health narratives.

3.4. Semantic Knowledge Graph Construction (f_4)

Once the observations have been synchronised, the transformation $f_4 : O' \to G$ converts the data into Resource Description Framework (RDF) triples using the COPE ontology. COPE models an individual's health as a collection of interconnected entities representing observations, activities, trajectories, temporal progression, and contextual relationships. Each RDF statement is expressed as a (s, p, o) triple (subject, predicate, object). The complete graph $G = (V, E)$ is structured with vertices V and directed edges E:

- Nodes ($V = P \cup A \cup O \cup Tr \cup L$): Represent Persons ($P$, e.g., `cope:Person`), Activities (A, e.g., `cope:Activity`), Observations (O, e.g., `cope:Observation`), Trajectories (Tr, e.g., `cope:Trajectory`), and Locations/Spatial Observations (L, e.g., `cope:SpatialObservation`).
- Edges ($E \subseteq V \times R \times V$): Represent semantic relationships where; $R = \{\text{performed}, \text{hasObservation}, \text{hasTrajectory}, \text{occurredAt}, \text{hasValue}, \text{hasUnit}\}$.

Unlike the original Apple Health export, where measurements exist as independent XML records, the resulting knowledge graph establishes explicit semantic relationships (e.g., `cope:performed`, `cope:hasObservation`, `cope:occurredAt`, `cope:value`, `cope:unit`, `cope:observedProperty`). This transforms isolated measurements into an interconnected graph that preserves temporal and contextual dependencies.

3.5. Semantic Enrichment (f_5)

Converting XML into RDF alone does not yield analytical value. Therefore, the framework performs semantic enrichment $f_5 : G \to G'$ by linking observations with contextual information derived across modalities. For an observation o_i, the framework constructs a contextual neighbourhood; $\Gamma(o_i) = \{a_i, g_i, l_i, o_i^{\text{adj}}\}$ where a_i is the activity session, g_i is the trajectory segment, l_i is the geographical location, and o_i^{adj} represents neighbouring physiological observations in time. Through this, a heart rate observation becomes explicitly connected to cadence, walking speed, distance, elevation changes, and energy expenditure. To facilitate interoperability with external biomedical terminologies, device-specific property identifiers (P_{device}) are mapped to clinical vocabularies (P_{clinical}) via a mapping, $\mu : P_{\text{device}} \to P_{\text{clinical}}$.

Device-specific identifiers are retained as provenance-bearing source identifiers and are mapped to semantic clinical concepts through ($\mu : P_{device} \to P_{clinical}$). For example, the Apple Health identifier HKQuantityTypeIdentifierHeartRate is mapped to the corresponding heart-rate concept (such as LOINC 8867-4). This distinction separates the original source identifier from its interoperable semantic meaning and allows both provenance and cross-source interoperability to be preserved. Queries therefore operate on the normalised semantic property rather than on an application-specific identifier.

3.6. Semantic Storytelling and Narrative Generation (f_6)

The principal objective of the framework is to support explainable, data-driven health storytelling. In this work, a health story is defined as a temporally ordered sequence of semantically related events together with the contextual evidence that explains the transition between those events. Thus, a story is not simply a visualisation of measurements; rather, the visualisations provide evidence for an interpretable account of what occurred, when it occurred, and which contextual observations support that interpretation.

Rather than visualising isolated measurements, the function $f_6 : G' \to S$ aggregates semantically related observations into a health story $S = (s_1, s_2, \ldots, s_r)$. Each story consists of ordered semantic events $s_i = (E_i, C_i, T_i)$, where E_i denotes the event type, C_i represents its contextual relationships, and T_i is its temporal interval. For instance, a sequence of observations satisfying specific rule conditions:

$$\text{Heart Rate} > 150 \text{ bpm}$$
$$\text{Speed} > 5 \text{ km/h}$$
$$\text{Cadence} > 150 \text{ steps/min}$$

is classified as an event $E_i = \text{HighIntensityExercise}$. Consequently, a workout session is interpreted as a chronological narrative sequence of health episodes:

$$\text{WarmUp} \rightarrow \text{ModerateExercise} \rightarrow \text{HighIntensityExercise} \rightarrow \text{Recovery}$$

These episodes emerge through temporal relationships among multiple concurrent observations. A period of increasing heart rate accompanied by increasing cadence, speed, and energy expenditure is interpreted as an intensification of physical effort, while stable gait symmetry provides contextual evidence supporting healthy movement.

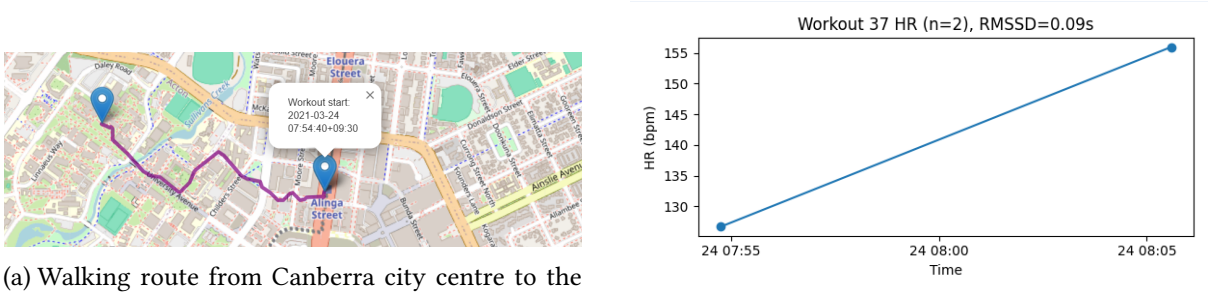
3.7. Visualisation and Semantic Querying (f_7)

To communicate these health narratives, the resulting graph can be inspected through graph visualisation and queried through SPARQL. Additionally, representing the data as a semantic graph G' enables expressive graph queries $Q = (V_Q, E_Q)$ where $Q \subseteq G'$. A query Q defined as:

$$Q = \{\text{HR} > 150, \text{Speed} > 5, \text{Cadence} > 150, \text{Energy} > 30\}$$

retrieves the subgraph matching set $R_Q = \{o_i \mid o_i \vDash Q\}$, isolating periods of sustained, vigorous activity. These retrieved segments can be visualised spatially or incorporated into automatically generated narrative text.

4. Application Case Study: Semantic Health Trajectory Construction



(a) Walking route from Canberra city centre to the Australian National University.

(b) Heart rate during the workout shown in Figure 2a.

Figure 2: A workout route and the corresponding heart rate during the activity.

To demonstrate the applicability of the proposed semantic modelling pipeline, we use longitudinal Apple Health data collected through an Apple Watch between 2017 and 2025. The dataset contains observations spanning multiple years for a single individual. The purpose of using this longitudinal source is to demonstrate that the framework can accommodate temporally distributed wearable observations; it is not intended to provide a population-level or longitudinal clinical analysis. For detailed evaluation, we select one representative workout session recorded on 24 March 2021 and use it as a controlled case study for examining semantic integration and querying. The source code is available online on GitHub[3].

[3]See: https://github.com/AsaraSenaratne/apple-health-data-modelling

```
<purl.archive.org/cope/record/57915>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierHeartRate";
cope:time "2021-03-24T07:54:45+09:30"^^xsd:dateTime;
cope:unit "count/min"; cope:value "126.648" .
```

(a) Heart rate variations

```
<purl.archive.org/cope/record/125607>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierDistanceWalkingRunning";
cope:time "2021-03-24T07:54:12+09:30"^^xsd:dateTime;
cope:unit "km"; cope:value "0.847309" .
```

(b) Workout distance

```
<purl.archive.org/cope/record/858583>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierStepCount";
cope:time "2021-03-24T07:54:12+09:30"^^xsd:dateTime;
cope:unit "count"; cope:value "952" .
```

(c) Step count and walking asymmetry

```
<purl.archive.org/cope/record/426929>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierAppleStandTime";
cope:time "2021-03-24T07:55:58+09:30"^^xsd:dateTime;
cope:unit "min"; cope:value "5" .
```

(d) Stand time

```
<purl.archive.org/cope/record/403704>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierActiveEnergyBurned";
cope:time "2021-03-24T08:04:48+09:30"^^xsd:dateTime;
cope:unit "kCal"; cope:value "39.532" .
```

(e) Energy burned and walking step length

```
<purl.archive.org/cope/record/437821>
a cope:Observation; cope:observedProperty
"HKQuantityTypeIdentifierWalkingSpeed";
cope:time "2021-03-24T07:58:31+09:30"^^xsd:dateTime;
cope:unit "km/h"; cope:value "5.004" .
```

(f) Walking speed

Figure 3: Extracts of Apple Health observations related to the workout in Figure 2a, mapped using the COPE ontology.

The analysis focuses on an individual workout session recorded on 24th March 2021. The original wearable export contains heterogeneous measurements, including cardiovascular signals, activity summaries, energy expenditure, mobility indicators, and spatial trajectories. These records are transformed into COPE-compliant observations and connected through temporal and contextual relationships. The resulting graph enables observations that originate from different sensor modalities to be queried jointly. This case study therefore evaluates the expressive capability of the semantic representation rather than predictive performance.

As shown in Figure 1, our pipeline for semantically modelling wearable-device data begins with raw data acquisition from two primary sources: (1) structured sensor export files provided by the wearable ecosystem, and (2) optional GPS Exchange Format (GPX) files or equivalent spatial logs generated by the wearable or its paired mobile device. During ingestion, sensor exports are parsed to extract measurement records, activity summaries, and device metadata. Timestamps are standardised and transformed into structured observation and activity tables. When available, spatial logs are parsed to retrieve GPS track points and segments, producing a trajectory table with associated temporal metadata. These outputs feed into a matching and linkage layer that integrates activity metadata, spatial traces, and sensor observations. Activities are temporally aligned with trajectories, and an interval-based matching procedure links sensor observations to activities within defined time windows. The integrated data are then transformed into a KG compliant with the COPE ontology [1]. Activities are represented as `cope:Activity`, observations as `cope:Observation`, and GPS-derived paths as `cope:SpatialObservation` and `cope:Trajectory`, all associated with a `cope:Person`. Semantic relations such as `cope:performed`, `cope:hasObservation`, `cope:unit`, and `cope:value` structure the graph. Finally, the visualisation and export layer produces user-facing artefacts such as Figures 2a and 2b.

In Figure 2, we illustrate the integration of spatial and physiological information for the selected workout. The GPS trajectory is represented as a semantic trajectory entity and temporally aligned with heart-rate observations. The resulting representation enables interpretation of physiological responses within the context of movement rather than analysing heart-rate measurements in isolation. During the initial phase of the activity, heart-rate observations remain around 129 bpm, corresponding to a warm-up period. As the activity progresses, the heart rate increases gradually, reaching approximately 145 bpm during the main exercise phase and a maximum of approximately 155 bpm. Distance, heart rate, and energy expenditure (39.5 kcal) together suggest a short, moderate-intensity workout. Gait measurements show a consistent walking asymmetry of 0%, indicating stable and symmetric movement, while stand-time observations confirm continuous upright activity consistent with the GPS trace.

Figure 3 presents examples of semantically modelled wearable observations generated from the

same activity. Each observation is represented with explicit properties describing the observed feature, timestamp, measurement unit, and value. For example, heart rate is represented as a cardiovascular observation, while distance, step count, energy expenditure, and walking speed are represented as corresponding COPE observation instances. This semantic representation allows observations from different sensing modalities to be queried and analysed together.

5. Evaluation

To evaluate the expressive capability of the proposed semantic representation, we formulate four competency questions.

1. What was the progression of heart rate during exercise periods?
   Goal: *Retrieve time-series heart rate observations linked to physical activity/exercise sessions, ordered chronologically to track progression.*

Listing 1: SPARQL query to retrieve time-series heart rate observations linked to physical activity/exercise sessions, ordered chronologically to track progression.

```
PREFIX cope: <https://purl.archive.org/cope#>>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>\
PREFIX rdf: <http://www.w3.org/1999/02/22-rdf-syntax-ns#>
PREFIX rdfs: <http://www.w3.org/2000/01/rdf-schema#>

SELECT ?workoutType ?timestamp ?hrValue ?unit
WHERE {
  ?workout a cope:Activity ;
      cope:activityType ?workoutType ;
      cope:startTime ?workoutStart ;
      cope:endTime ?workoutEnd .
  ?obs a cope:Observation ;
      cope:observedProperty "HKQuantityTypeIdentifierHeartRate" ;
      cope:value ?hrValue ;
      cope:unit ?unit ;
      cope:time ?timestamp .
  FILTER (?timestamp >= ?workoutStart && ?timestamp <= ?workoutEnd)
}
ORDER BY ?workoutType ?timestamp
```

2. Which activities resulted in elevated cardiovascular response? Goal: *Identify specific activity types where average or peak heart rate exceeded a defined threshold (such as > 140 BPM).*

Listing 2: SPARQL query to retrieve specific activity types where average or peak heart rate exceeded a defined threshold (such as > 140 BPM).

```
PREFIX cope: <https://purl.archive.org/cope#>>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?activityType (AVG(?hrValue) AS ?avgHR) (MAX(?hrValue) AS ?peakHR)
WHERE {
  ?activity a cope:Activity ;
       cope:activityType ?activityType ;
       cope:startTime ?actStart ;
       cope:endTime ?actEnd .

  ?obs a cope:Observation ;
      cope:observedProperty "HKQuantityTypeIdentifierHeartRate" ;
      cope:value ?hrValue ;
      cope:time ?timestamp .

  FILTER (?timestamp >= ?actStart && ?timestamp <= ?actEnd)
}
GROUP BY ?activityType
HAVING (AVG(?hrValue) > 140) # Adjust threshold for elevated response
ORDER BY DESC(?peakHR)
```

3. How did physical activity patterns change over time? Goal: *Aggregate workout frequency, duration, or step progression over temporal trajectory windows (such as weekly or monthly evolution).*

Listing 3: SPARQL query to aggregate workout frequency, duration, or step progression over temporal trajectory windows (such as weekly or monthly evolution).

```
PREFIX cope: <https://purl.archive.org/cope#>>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?trajectoryWindow (COUNT(?activity) AS ?totalSessions) (SUM(?duration) AS ?totalDurationMinutes) (SUM(?steps) AS ?totalSteps)
WHERE {
 ?trajectory a cope:HealthTrajectory ;
          cope:timeWindow ?trajectoryWindow ;
          cope:hasProgressionEvent ?event .

 ?event cope:associatedActivity ?activity .

 ?activity cope:durationInMinutes ?duration ;
         cope:stepCount ?steps .
}
GROUP BY ?trajectoryWindow
ORDER BY ?trajectoryWindow
```

4. Identify periods with abnormal physiological response Goal: *Detect trajectory windows or specific timeframes where physiological metrics (such as Resting Heart Rate or Exercise Heart Rate) deviated significantly from baseline levels.*

Listing 4: SPARQL query to retrieve detected trajectory windows or specific timeframes where physiological metrics (such as Resting Heart Rate or Exercise Heart Rate) deviated significantly from baseline levels.

```
PREFIX cope: <https://purl.archive.org/cope#>>
PREFIX xsd: <http://www.w3.org/2001/XMLSchema#>

SELECT ?timestamp ?hrValue
     ("Elevated Resting HR" AS ?anomalyType)
WHERE {
   ?obs a cope:Observation ;
       cope:observedProperty "RestingHeartRate" ;
       cope:value ?hrValue ;
       cope:time ?timestamp .

   BIND(65.0 AS ?baselineValue) #baseline of 65 BPM

   FILTER (?hrValue > (?baselineValue * 1.20))
}
ORDER BY DESC(?timestamp)
```

6. Conclusion and Future Work

This paper presented a semantic modelling approach that transforms heterogeneous wearable sensor observations into structured, temporally connected personal health trajectories. By representing physiological observations, activities, trajectories, and contextual relationships within a COPE-based knowledge graph, the approach moves beyond isolated measurements and application-specific visualisations towards an interoperable representation that can be queried and reused for data-driven health storytelling.

Using longitudinal Apple Health data as the source and a representative workout as the case study, we demonstrated the ability of the resulting semantic representation to answer four competency questions covering temporal physiological progression, elevated cardiovascular response, activity progression, and deviations from a physiological baseline. The results demonstrate semantic expressiveness and cross-modal queryability rather than clinical validity or the effectiveness of a predictive model. More broadly, the work highlights how explicit semantic relationships can provide the evidential foundation for narrative representations of personal health data.

Future work will extend the framework in three directions: (1) automated semantic reasoning and anomaly detection using SHACL constraints and OWL 2 RL reasoning; (2) predictive trajectory analytics using graph neural networks and embedding models; and (3) privacy-preserving federated knowledge graphs that allow individuals to retain control over their personal health trajectory data while supporting secure health research. A further priority is evaluating the understandability and usefulness of generated health stories through user studies and comparison with conventional fitness-platform representations.

Declaration on Generative AI

During the preparation of this work, the authors used Grammarly solely for grammar checking and language refinement. After using this tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] A. Senaratne, O. Seneviratne, H. Z. Lim, L. Seneviratne, Cope: Chronic observation and progression events ontology, in: CEUR, volume 4065, 2025, pp. 72–86.

[2] A. Senaratne, Anomaly detection in graphs for knowledge discovery and data quality enhancement, Open Research Repository-Australian National University (2024).

[3] A. Senaratne, P. Christen, G. Williams, P. G. Omran, Rule-based knowledge discovery via anomaly detection in tabular data, in: CEUR Workshop Proceedings, volume 3433, CEUR-WS, 2023.

[4] T. T. N. Nguyen, A. Senaratne, L. Seneviratne, Etcod: embedding-based anomaly detection and llm-driven validation framework for knowledge graphs, in: Australasian Conference on Data Science and Machine Learning, Springer, 2025, pp. 454–469.

[5] G. Hripcsak, J. D. Duke, N. H. Shah, C. G. Reich, V. Huser, M. J. Schuemie, M. A. Suchard, R. W. Park, I. C. K. Wong, P. R. Rijnbeek, et al., Observational health data sciences and informatics (ohdsi): opportunities for observational researchers, Studies in health technology and informatics 216 (2015) 574.

[6] P. Chandak, K. Huang, M. Zitnik, Building a knowledge graph to enable precision medicine, Scientific data 10 (2023) 67.

[7] A. Senaratne, Seka: Seeking knowledge graph anomalies, in: Web Conference, 2023, pp. 568–572.

[8] A. Senaratne, P. Christen, G. Williams, P. G. Omran, Unsupervised identification of abnormal nodes and edges in graphs, ACM Journal of Data and Information Quality 15 (2022) 1–37.

[9] A. Senaratne, P. G. Omran, G. Williams, P. Christen, Unsupervised anomaly detection in knowledge graphs, in: Proceedings of the 10th international joint conference on knowledge graphs, 2021, pp. 161–165.

[10] H. K. Rudsari, B. Tseng, H. Zhu, L. Song, C. Gu, A. Roy, E. Irajizad, J. Butner, J. Long, K.-A. Do, Digital twins in healthcare: a comprehensive review and future directions, Frontiers in Digital Health 7 (2025) 1633539.

[11] Y. Lu, C. Liu, I. Kevin, K. Wang, H. Huang, X. Xu, Digital twin-driven smart manufacturing: Connotation, reference model, applications and research issues, Robotics and computer-integrated manufacturing 61 (2020) 101837.

[12] M. Compton, P. Barnaghi, L. Bermudez, R. Garcia-Castro, O. Corcho, S. Cox, J. Graybeal, M. Hauswirth, C. Henson, A. Herzog, et al., The ssn ontology of the w3c semantic sensor network incubator group, Journal of Web Semantics 17 (2012) 25–32.

[13] K. Janowicz, A. Haller, S. J. Cox, D. Le Phuoc, M. Lefrançois, Sosa: A lightweight ontology for sensors, observations, samples, and actuators, Journal of Web Semantics 56 (2019) 1–10.

[14] M. Bermudez-Edo, T. Elsaleh, P. Barnaghi, K. Taylor, Iot-lite: a lightweight semantic model for the internet of things, in: 2016 INTL IEEE conferences on ubiquitous intelligence & computing, advanced and trusted computing, scalable computing and communications, cloud and big data computing, internet of people, and smart world congress (uic/atc/scalcom/cbdcom/iop/smartworld), IEEE, 2016, pp. 90–97.

[15] R. García-Castro, M. Lefrançois, M. Poveda-Villalón, L. Daniele, The etsi saref ontology for smart applications: a long path of development and evolution, Energy smart appliances: applications, methodologies, and challenges (2023) 183–215.

[16] J. Moreira, L. F. Pires, M. Van Sinderen, L. Daniele, M. Girod-Genet, Saref4health: Towards iot standard-based ontology-driven cardiac e-health systems, Applied ontology 15 (2020) 385–410.

[17] M. Botts, A. Robin, Opengis® sensor model language (sensorml) implementation specification. version 1.0. 0. (2007).
[18] M. Alrehiely, P. Eslambolchilar, R. Borgo, A taxonomy for visualisations of personal physical activity data on self-tracking devices and their applications, in: Proceedings of the 32nd International BCS Human Computer Interaction Conference, BCS Learning & Development, 2018.
[19] G. Chan, C. Nwagu, I. Odenigbo, A. Alslaity, R. Orji, The shape of mobile health: a systematic review of health visualization on mobile devices, International Journal of Human–Computer Interaction 41 (2025) 1154–1172.
[20] R. Reda, F. Piccinini, G. Martinelli, A. Carbonaro, Heterogeneous self-tracked health and fitness data integration and sharing according to a linked open data approach, Computing 104 (2022) 835–857.
[21] A. Gyrard, M. Gaur, S. Shekarpour, K. Thirunarayan, A. Sheth, Personalized health knowledge graph, in: CEUR workshop proceedings, volume 2317, 2018, p. 5.
[22] C. N. Vorisek, M. Lehne, S. A. I. Klopfenstein, P. J. Mayer, A. Bartschke, T. Haese, S. Thun, Fast healthcare interoperability resources (fhir) for interoperability in health research: systematic review, JMIR medical informatics 10 (2022) e35724.
[23] N. Beredimas, V. Kilintzis, I. Chouvarda, N. Maglaveras, A reusable ontology for primitive and complex hl7 fhir data types, in: 2015 37th Annual International Conference of the IEEE Engineering in Medicine and Biology Society (EMBC), IEEE, 2015, pp. 2547–2550.
[24] K. A. Spackman, K. E. Campbell, R. A. Côté, Snomed rt: a reference terminology for health care, in: Proceedings of the AMIA annual fall symposium, 1997, p. 640.
[25] C. J. McDonald, S. M. Huff, J. G. Suico, G. Hill, D. Leavelle, R. Aller, A. Forrey, K. Mercer, G. DeMoor, J. Hook, et al., Loinc, a universal standard for identifying laboratory observations: a 5-year update, Clinical chemistry 49 (2003) 624–633.
[26] B. Smith, M. Ashburner, C. Rosse, J. Bard, W. Bug, W. Ceusters, L. J. Goldberg, K. Eilbeck, A. Ireland, C. J. Mungall, et al., The obo foundry: coordinated evolution of ontologies to support biomedical data integration, Nature biotechnology 25 (2007) 1251–1255.
[27] D. S. Himmelstein, A. Lizee, C. Hessler, L. Brueggeman, S. L. Chen, D. Hadley, A. Green, P. Khankhanian, S. E. Baranzini, Systematic integration of biomedical knowledge prioritizes drugs for repurposing, Technical Report, Manubot, 2017.
[28] K. Kaur, M. Selway, A. Senaratne, M. Stumptner, G. Grossmann, Interoperable digital twin ecosystems: enhancing risk management in distributed energy systems, Australian Energy Producers Journal 66 (2026) EP25146.
[29] A. Senaratne, P. Christen, P. Omran, G. Williams, Taxonomy of anomaly types in knowledge graphs, in: ESWC, Springer, 2025, pp. 151–155.
[30] H. Kim, J. Jung, J. Choi, Developing a dietary lifestyle ontology to improve the interoperability of dietary data, JMIR 6 (2022).
[31] M. Dragoni, T. Bailoni, et al., Integrating functional status information into knowledge graphs to support self-health management, Data Intelligence 5 (2023).
[32] C. J. McDonald, et al., Loinc, a universal standard for identifying laboratory observations, Clinical chemistry 49 (2003) 624–633.
[33] S. Schulz, G. O. Klein, Snomed ct–advances in concept mapping, retrieval, and ontological foundations, BMC MIDM 8 (2008) S1.