www.itcon.org - Journal of Information Technology in Construction - ISSN 1874-4753

TOWARDS A SEMANTIC DIGITAL TWIN FOR CONSTRUCTION SITE PM EMISSION MANAGEMENT: THRESHOLD MODELLING AND VALIDATION FOR ALERTING WORKFLOWS

SUBMITTED: June 2026
PUBLISHED: August 2026
EDITOR: Farzad Rahimian
DOI: 10.36680/j.itcon.2026.041

Alessandro Bruttini, Ph.D.
Department of Architecture, University of Florence, Italy
https://orcid.org/0000-0001-6317-2406
alessandro.bruttini@unifi.it

Philipp Hagedorn, Ph.D.
Chair of Computing in Engineering, Ruhr University Bochum, Germany
https://orcid.org/0000-0002-6249-243X
philipp.hagedorn-n6v@rub.de

Vito Getuli, Ph.D.
Department of Civil, Building and Environmental Engineering, Sapienza University of Rome, Italy
https://orcid.org/0000-0001-8470-2648
vito.getuli@uniroma1.it

Tommaso Sorbi, Ph.D.
Department of Architecture, University of Florence, Italy
https://orcid.org/0000-0002-7204-0437
tommaso.sorbi@unifi.it

Pietro Capone, Professor
Department of Architecture, University of Florence, Italy
https://orcid.org/0000-0001-9294-7653
pietro.capone@unifi.it

SUMMARY: *Particulate matter (PM) emissions from construction activities affect urban air quality, worker exposure, and surrounding communities. Improving their management is therefore a relevant contribution to sustainability and reflects UN Sustainable Development Goals (SDGs) related to health and well-being (SDG 3), sustainable infrastructure (SDG 9), and urban liveability (SDG 11). Digital Twins (DTs) can support PM emission management by connecting monitoring data, site context, threshold information, and decision-support services. However, existing approaches still lack structured semantic workflows able to represent threshold entities, evaluate PM observations, and retrieve alert-relevant context. Building on a previous high-level DT framework for PM emission management on urban construction sites and on the Construction Site Emission Management Ontology (XEMO), this paper proposes and evaluates a semantic alerting workflow based on threshold modelling and SHACL validation. Thresholds derived from current and forthcoming European environmental air quality reference values are represented as semantic entities linked to pollutants, averaging periods, units, values, and regulatory sources. A demonstrator is then developed in which synthetic PM10 and PM2.5 observations are generated, stored as time-series data, transformed into semantic data according to the Resource Description Framework (RDF), integrated into a knowledge graph, and validated against threshold-based warning conditions through SHACL shapes. The validation reports are queried to retrieve the related alert context, including pollutant, measured concentration, threshold condition, sensor, construction area, and emission source. The results show the prototype-level feasibility of using semantic modelling and SHACL validation to support traceable PM alerting, providing a reusable basis for future decision-support services within semantic DTs and real-site implementation.*

KEYWORDS: *semantic Digital Twin, construction site, PM Emission, SHACL, decision support.*

REFERENCE: *Bruttini, A., Hagedorn, P., Getuli, V., Sorbi, T., & Capone, P. (2026). Towards a semantic digital twin for construction site PM emission management: Threshold modelling and validation for alerting workflows. Journal of Information Technology in Construction (ITcon), 31, 970-989. https://doi.org/10.36680/j.itcon.2026.041*

1. INTRODUCTION

Particulate matter (PM) emissions generated by construction activities remain a relevant challenge for urban environmental quality, public health, and occupational safety. This issue is particularly significant in dense urban areas, where construction sites intensify already critical pollution conditions and affect residents, public spaces, and sensitive facilities in their immediate surroundings. Improving the monitoring and management of these emissions is therefore connected not only to site-level environmental control, but also to broader sustainability objectives, including the protection of health and well-being, the development of more sustainable and innovative infrastructure, and the improvement of urban liveability, as reflected in the United Nations Sustainable Development Goals (SDG 3, SDG 9, and SDG 11; (United Nations, 2015a, 2015b, 2015c)).

Construction-related PM emissions are associated with several recurring activities, including excavation, demolition, earthworks, material handling, and vehicle circulation. Their impacts on air quality concern both workers directly exposed on site and nearby communities affected by the dispersion of pollutants beyond site boundaries (Azarmi et al., 2016; Cheriyan & Choi, 2020; C. Z. Li et al., 2019). Moreover, monitoring and mitigating these emissions remains particularly challenging because construction sites are inherently complex and continuously changing environments, where work areas, activities, equipment, mitigation measures, and exposure conditions vary over time. As a result, effective monitoring approaches must support not only PM data collection, but also their contextual interpretation against evolving site conditions and the timely activation of mitigation responses.

To address this need, recent research has increasingly explored digital and data-driven solutions for PM monitoring, especially through sensor networks and IoT-based observation systems. However, these developments still appear fragmented, as individual technological components are often investigated separately and rarely integrated into systems capable of supporting monitoring, alerting, and decision-making across multiple stakeholders and operational conditions (Bruttini, Sorbi, et al., 2026b). This challenge is further compounded by the lack of consolidated reference thresholds for occupational exposure assessment and alerting, since existing criteria are often derived from environmental air quality standards and are not always suited to the short-term and task-specific exposure conditions that characterize construction sites.

Against this background, the Digital Twin (DT) paradigm offers a suitable framework for structuring construction site PM emission management around the continuous connection between the physical site, its digital representation, and the services required to support monitoring and decision-making. This paper builds on a previous contribution by the authors, which proposed a high-level DT framework for PM emission monitoring and management on urban construction sites (Bruttini, Sorbi, et al., 2026a). The framework identifies the main stakeholders, data sources, and use cases: monitoring and visualisation, alerting, and planning. The present study uses this framework as the basis for progressing from the conceptual definition of requirements towards system operationalisation. More specifically, it focuses on how threshold information and PM observations can be semantically represented, integrated, and validated to support the alerting function within a future DT system.

A related contribution by the authors developed the Construction Site Emission Management Ontology (XEMO) for representing construction site organization, emission sources, pollutants, observations, thresholds, and their relations (Bruttini, Hagedorn, et al., 2026). Building on XEMO, this paper extends the previous DT framework through the further development and implementation of the demonstrator system used for the preliminary evaluation of the ontology capabilities. In the demonstrator, synthetic PM10 and PM2.5 observations are generated and stored as time-series data, transformed into semantic data using the Resource Description Framework (RDF), integrated into a knowledge graph, and evaluated against modelled alert thresholds through Shapes Constraint Language (SHACL) validation. Through this operational step, the paper further articulates the modelling of threshold entities and alert-triggering logics, showing how semantically integrated PM observations can be validated against explicit threshold conditions to generate alert-relevant outputs, where each detected exceedance is connected to its pollutant, measured value, threshold value, sensor, construction area, and related emission source. This lays the basis for future DT-based services such as dashboard notifications, compliance logs, mitigation workflows, and real-site testing with sensor data.

The remainder of the paper follows with Section 2 providing the background on DTs, semantic technologies, and PM pollution management in construction. Section 3 presents the research framework and methodology. Section

4 describes threshold modelling and alerting logic. Section 5 presents the demonstrator and validation workflow. Section 6 discusses the results, current limitations, and future work. Section 7 draws the main conclusions.

2. BACKGROUND

This section reviews the background needed to frame the proposed approach. It first discusses digital twin developments in the Architecture, engineering, Construction and Operation (AECO) sector, focusing on semantic integration and modular architectures. It then examines existing digital and ontology-based approaches for PM pollution management in construction, leading to the research gaps addressed by the proposed threshold modelling and validation workflow.

2.1 Digital Twins and semantic integration in the AECO sector

The digital twin concept originated in the manufacturing domain, in connection with product lifecycle management, and describes the association between a physical entity, its digital counterpart, and the data connection that keeps the two aligned over time (Grieves, 2005; F. Tao et al., 2019). In the AECO sector, DTs have developed in close connection with Building Information Modelling (BIM), since BIM provides structured asset information that can be enriched with operational, environmental, and process-related data (Borrmann et al., 2018). This interpretation has progressively expanded towards data-centric systems that integrate heterogeneous static and dynamic data from models, sensors, repositories, and processing services to support monitoring, diagnosis, prediction, and decision-making throughout the project lifecycle (Boje et al., 2022; Sacks et al., 2020).

In the construction phase, this integration problem is amplified by the temporary and changing nature of sites, where spaces, activities, resources, equipment, and environmental conditions evolve during project execution. For this reason, recent DT research has emphasised architectures able to connect project information with operational and environmental data across different spatial and temporal scales (Boje et al., 2022; Han et al., 2025; Sacks et al., 2020). This perspective also extends to urban-scale DTs, where building, infrastructure, geospatial, and city-level datasets are combined to support wider monitoring and management functions (Adreani et al., 2024; Corneli & Rotilio, 2023; Weil et al., 2023).

Semantic Web technologies address this integration need through structured data modelling and machine-interpretable representation. Ontologies and Linked Data principles (Berners-Lee, 2006) allow heterogeneous datasets to be described through shared identifiers, explicit relations, and reusable vocabularies. In the AECO sector, this direction is supported by the Linked Building Data (LBD) approach and by a growing ecosystem of ontologies for the description of spatial entities, sensing and observations, temporal information, provenance, and construction processes (Farghaly et al., 2023; W3C Linked Building Data Community Group, 2022). Recent contributions have further extended this landscape through ontology frameworks, modular Semantic Digital Twin architectures, and System-of-Systems approaches for construction DTs (Hagedorn et al., 2025; Kosse et al., 2025; Schlenger et al., 2025). These developments support interoperability, traceability, knowledge extraction, and reasoning over datasets produced by different tools and stakeholders.

Semantic reasoning and rule-based checking have also been applied across various domains including construction safety and regulatory compliance. Lu et al. (2015) developed CSCOntology to represent construction tasks, hazard precursors, hazards, and preventive measures, together with safety constraints expressed through SWRL rules. The approach was evaluated using a demonstrative dataset representing a static roofing safety scenario, without integrating dynamic data. Jiang et al. (2022) transformed IFC-based building models into a design-model ontology and combined them with code and compliance-checking ontologies through mapping and checking rules to assess static building designs. The checking results were presented as reports intended for consultation, limiting their reuse in subsequent semantic processing and querying. Li et al. (2022) integrated static BIM information with real sensor measurements represented as RDF instances and applied automatically generated SPARQL rules to identify safety risk factors that did not satisfy regulatory constraints in subway construction. The approach supported dynamic safety checking, although its evaluation used a one-day historical monitoring dataset rather than a continuously operating data stream. Taken together, these studies show an evolution from ontology-based representation and rule checking of static scenarios towards the semantic integration and automated evaluation of dynamic monitoring data. These capabilities provide important foundations for Semantic Digital Twins by enabling contextualised reasoning over heterogeneous and evolving information. However, the reviewed approaches remain largely task-specific and demonstrative. Their integration into mature operational SDT architectures, and their broader application to continuous monitoring, alerting, and decision support, remain only partially developed.

2.2 Digital Twins and ontologies for PM pollution management in construction

Research on particulate matter emission management in construction has mainly addressed how PM concentrations can be measured, interpreted, mitigated, and predicted under site conditions. Monitoring studies commonly rely on PM10 and PM2.5 sensors, low-cost optical particle counters, and IoT acquisition systems to collect concentration data during construction operations (Cheriyan & Choi, 2020; Khan et al., 2021; Kim et al., 2021; Paolucci et al., 2022). These solutions provide frequent and spatially distributed PM concentration data, although their deployment on construction sites requires attention to calibration, dust concentration ranges, weather sensitivity, sensor robustness, and data reliability (Brostrøm et al., 2025; Jaafar et al., 2024; Zheng et al., 2024). Integrated hardware and software systems have also been proposed to support real-time environmental monitoring of dust and related pollutants, including noise and vibration, extending the monitoring scope towards broader site-level environmental control (Kang et al., 2021; Patel et al., 2024).

Contextual and environmental variables, including weather conditions, background pollution, activity type, equipment, workforce, material handling, and mitigation measures, have been used to interpret PM concentration patterns and relate observed values to site operations (Fang et al., 2025; Hong et al., 2020; Milivojević et al., 2023). Mitigation-oriented studies have examined dust suppression measures, including water spraying, showing the need to evaluate their effectiveness in relation to task type, local conditions, and monitored concentration trends (Brostrøm et al., 2025; Ham et al., 2024; Luo et al., 2021). Other contributions have explored decision-support functions through BIM-based layout optimisation, blockchain-based environmental data traceability, and worker-centred exposure alerting based on PM sensing and position data (Cho et al., 2021; Kim et al., 2025; G. Tao et al., 2022; Zhong et al., 2022).

Predictive approaches have used regression models, artificial neural networks, LSTM models, emission-factor methods, dispersion models, CFD reconstruction, and hybrid machine-learning workflows to estimate or forecast PM concentrations from sensor, meteorological, operational, and spatial data (Gill et al., 2025; Guo et al., 2022; Milivojevic et al., 2024; Wang et al., 2026; Y. Xu et al., 2025; Yoon et al., 2023). Recent multimodal approaches further integrate environmental sensing, video-based activity recognition, and historical PM data, supporting richer descriptions of the operational conditions associated with emission generation (Zhang et al., 2026). These contributions provide capabilities relevant to future DT systems, including forecasting, alerting, and proactive site control.

More explicit DT-oriented work for construction pollution management has recently emerged. Pham et al. developed a common data environment for a BIM-based DT to support real-time construction pollution management in building renovation projects, focusing on pollution data management and information sharing among project stakeholders (Pham et al., 2025). Subsequent work proposed a BIM-enhanced DT architecture for construction pollution management in building renovation projects, addressing the integration of BIM-based representation and pollution monitoring data within a DT-oriented platform (Pham et al., 2026). These studies provide relevant references for DT-oriented construction pollution management, particularly for BIM-based representation, real-time pollution data management, and stakeholder information exchange, while leaving outside their scope the semantic modelling of PM threshold entities, ontology-based processing of PM observations, and alerting logic.

The semantic layer required for these functions has been addressed only partially in adjacent areas of environmental and construction knowledge modelling. Zhong et al. (2018) proposed an ontology-based framework for building environmental monitoring and compliance checking under a BIM environment. Xu et al. (2022) developed an ontology model for construction site pollutant monitoring indicators. Lu et al. (2024) proposed the Construction Carbon Emission Ontology for carbon emission tracking and evaluation. These contributions demonstrate the relevance of formal semantic models for environmental information management, while their scopes differ from construction site PM alerting based on monitored pollutant observations and threshold validation. XEMO was developed to address this modelling need by representing construction site organisation, emission sources, pollutants, sensors, observations, thresholds, and their relations for PM-oriented alerting (Bruttini, Hagedorn, et al., 2026). Its initial evaluation demonstrated threshold exceedance detection and traceability capabilities, while further validation is required for the more detailed alerting logics investigated in this work and, subsequently, in workflows based on real site monitoring data.

2.3 Research gaps

The background outlined above identifies three gaps addressed by this paper:

- PM emission management in construction lacks structured DT workflows connecting monitoring data, site context, threshold information, and alert outputs within a coherent processing chain;
- semantic representation remains incomplete, as existing AECO and sensing ontologies require further specification to support threshold entities and alert-triggering conditions for construction site PM emission management;
- PM threshold logic for construction site alerting remains difficult to formalise, since available reference values mainly derive from environmental air-quality regulations and do not directly address short-term, task-specific, and location-dependent worker exposure.

Building on the previously developed DT framework for PM emission management in urban construction sites (Bruttini, Sorbi, et al., 2026a) and on XEMO as a domain ontology for construction site emission management (Bruttini, Hagedorn, et al., 2026), this paper addresses these gaps by focusing on threshold modelling and SHACL-based validation as a semantic workflow for detecting PM alert conditions and retrieving their relevant site context within a future DT system.

3. RESEARCH FRAMEWORK AND METHODOLOGY

The research gaps identified above indicate the need for a semantic DT-oriented workflow in which PM monitoring data, construction site context, threshold information, and alert outputs are connected through explicit threshold entities and alert-triggering conditions. This paper addresses this need as an implementation-oriented extension of the previously proposed DT framework for PM emission management on urban construction sites (Bruttini, Sorbi, et al., 2026a). The original framework identified the urban construction site as the physical system, the DT as the digital infrastructure for data storage, semantic integration, processing, and service delivery, and the relevant stakeholders as the users and beneficiaries of DT-supported PM management. Within this framework, three main DT services were identified: visualisation and monitoring, alerting, and planning. The present paper focuses on the alerting service, with specific attention to the semantic representation of threshold information, the integration of processed PM observations with site context, and the validation of warning conditions through SHACL. Figure 1 reports the reference framework and highlights the components addressed in this work, namely data acquisition, data integration and processing, and the PM threshold exceedance alerting service.

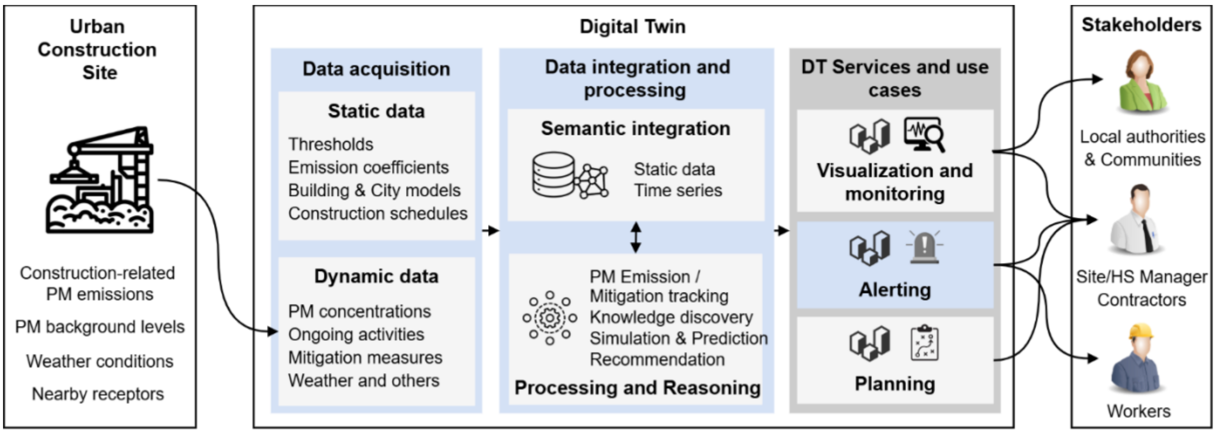

Figure 1: Conceptual framework of the Digital Twin for PM emission management on urban construction sites, adapted from Bruttini, Sorbi, et al. (2026a), with the components addressed in this paper highlighted in blue.

The semantic basis for this extension is provided by the Construction Site Emission Management Ontology (XEMO), developed by the authors in a related contribution that belongs to the same overarching research direction (Bruttini, Hagedorn, et al., 2026). XEMO follows a Linked Data approach to support the semantic integration of heterogeneous information required for construction site PM emission management. It reuses established ontologies where consolidated modelling patterns already exist: the Building Topology Ontology (BOT) (Rasmussen et al., 2021) is used to describe the spatial structure of the construction site, the Semantic Sensor Network Ontology (SSN/SOSA) (Haller et al., 2017) to represent sensors and observations, Quantities, Units, Dimensions and Types (QUDT) (FAIRsharing Team, 2015) to describe measured quantities and units of measure, OWL-Time (Cox & Little, 2022) to represent temporal entities and averaging periods, and PROV-O (Lebo et al., 2013) to link threshold entities to their regulatory sources. These reused ontologies are complemented by original XEMO classes and properties needed to cover domain-specific concepts that are not directly addressed by general AECO or sensing ontologies, including PM emission sources, pollutants, threshold entities, threshold kinds, alerting-related relations, and the links between construction areas, activities, sensors, observations, and emission sources. In the present paper, XEMO is used as the semantic backbone for modelling threshold entities, linking them to pollutants and processed PM observations, and supporting the SHACL-based validation workflow. Figure 2 provides a representation of the core XEMO modelling patterns used in this work, showing how the spatial representation of the construction site is connected to PM emission sources, observations, and threshold entities.

The methodology adopted in this paper comprises three main steps. The first step concerns PM threshold modelling in XEMO. Current and forthcoming European environmental air quality reference values are analysed and selected as provisional operational references for alerting, then represented as semantic threshold entities linked to pollutants, averaging periods, units, values, threshold kinds, and regulatory sources. The modelling choices and alerting logic are presented in Section 4.

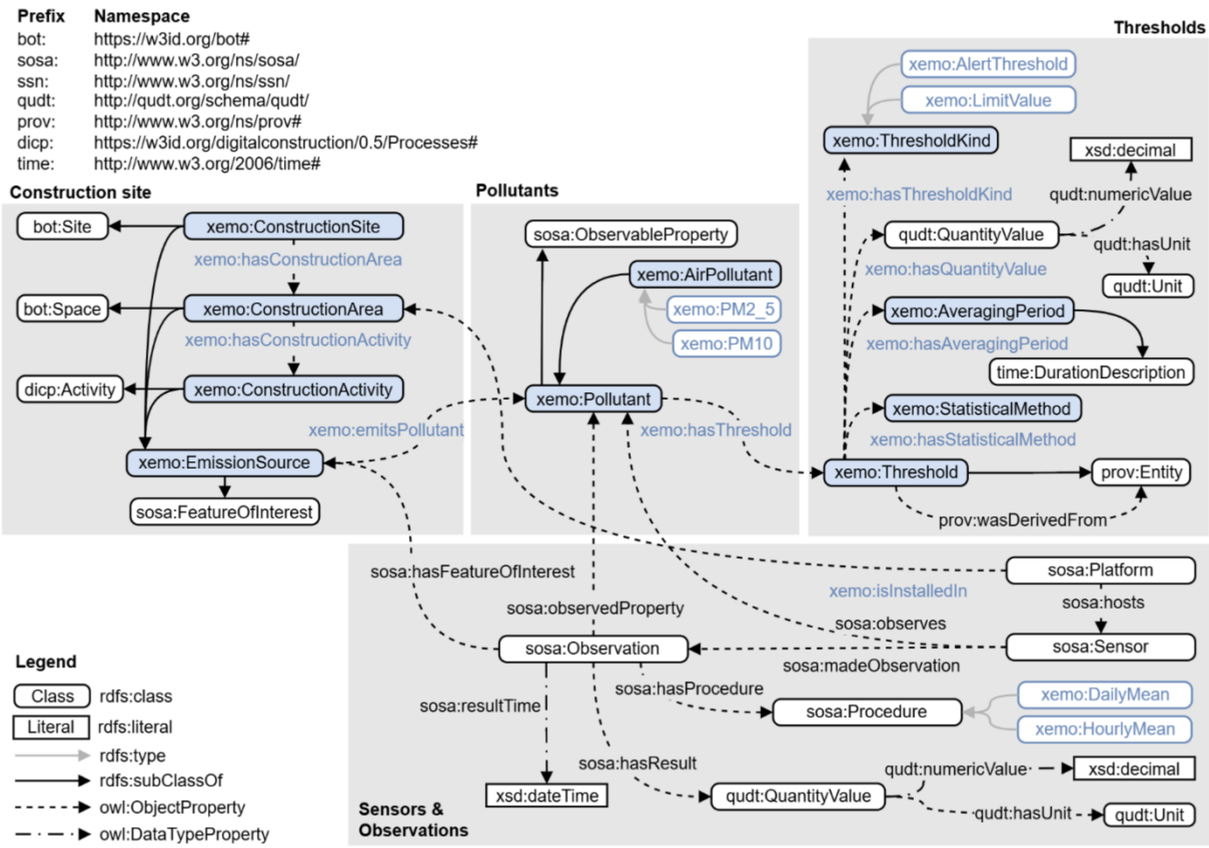

Figure 2: Graph representation of the XEMO ontology excerpt used for PM threshold modelling and alerting in this paper, adapted from Bruttini, Hagedorn, et al. (2026). Prefixes and namespaces of the reused ontologies are included; original XEMO classes and properties, and individuals are shown in blue.

The second step concerns the further development of the demonstrator system previously implemented for the XEMO capability assessment. In this paper, the demonstrator is extended towards the evaluation of an alerting workflow fitting the proposed DT framework. Synthetic PM10 and PM2.5 observations are generated, stored as time-series data, processed into rolling indicators, transformed into RDF, and integrated into a knowledge graph together with site and threshold information. The demonstrator architecture and implementation are described in Section 5.

The third step concerns SHACL-based alerting validation. SHACL shapes are defined to evaluate processed PM observations against the threshold-based warning logic proposed in this paper. The validation reports generated by this process are then queried through SPARQL to retrieve the corresponding alert context, including pollutant, measured concentration, threshold condition, sensor, construction area, and associated emission source. This step is used to assess the prototype-level feasibility of the alerting service within the proposed semantic DT direction.

4. THRESHOLDS MODELLING AND ALERTING LOGICS

4.1 Regulatory reference and threshold selection

In the proposed DT framework, thresholds provide the formal reference against which PM observations are evaluated to support the alerting service. In line with the scope of the research, the Italian and European regulatory frameworks were examined, showing that they do not currently provide dedicated occupational exposure limits specifically defined for PM10 and PM2.5 in occupational settings in general, nor for construction sites in particular. For this reason, ambient air quality standards are adopted in the development and testing of the site PM emission management demonstrator as an operational reference. These values are not equivalent to worker-specific exposure limits and are therefore interpreted in this study as provisional alerting criteria for site management.

Within the European framework, Directive (EU) 2024/2881 on ambient air quality and cleaner air for Europe revises the previous regulatory setting established by Directive 2008/50/EC, introducing stricter air quality standards for the protection of human health (European Parliament and Council, 2008, 2024). In the Italian context, the currently applicable reference remains the national implementation of Directive 2008/50/EC, while Directive (EU) 2024/2881 must be transposed into national regulations by 11 December 2026 and introduces a transitional phase before the stricter regime becomes applicable from 1 January 2030. Table 1 summarises the current and forthcoming reference values for PM10 and PM2.5. Compared with the current framework, the 2030 regime lowers the daily and annual limit values for both pollutants, reduces the number of allowed exceedances for daily limits per calendar year, and introduces PM information and alert thresholds intended to activate public information and short-term response measures when high concentration levels are measured or forecast.

Table 1: Current and forthcoming ambient air quality reference values for PM10 and PM2.5 thresholds under the Italian and European regulatory frameworks.

Pollutant	Averaging period	Current and transitional thresholds (to 01/01/30)	Future thresholds (from 01/01/30)	Information / alert thresholds (from 01/01/30)
PM10	Daily	50 μg/m³ (max 35/y exc.)	45 μg/m³ (max 18/y exc.)	90 μg/m³
	Annual	40 μg/m³	20 μg/m³	-
PM2.5	Daily	-	25 μg/m³ (max 18/y exc.)	50 μg/m³
	Annual	25 μg/m³	10 μg/m³	-

In this study, the 2030 values are adopted as the reference for modelling PM threshold entities and defining the alert-triggering logic, since they represent the forthcoming stricter regulatory regime. However, these thresholds are still defined for ambient air quality assessment and are based on daily averaging periods. Therefore, in the demonstrator, they are not used as direct occupational exposure limits, but as regulatory reference values from which operational alerting conditions for site management are derived.

4.2 Threshold modelling in XEMO

In XEMO, thresholds are represented as instances of the `xemo:Threshold` class. Each threshold is described through a set of properties and relationships that make explicit both its quantitative specification and its regulatory meaning. In particular, each threshold is linked to the pollutant to which it applies and specifies its quantity value, unit of measure, and threshold kind. The latter distinguishes limit values for annual and daily means from alert thresholds for information and alert conditions. Each threshold also specifies the averaging period and statistical aggregation method required for comparison with pollutant observations, together with its regulatory source. As a result, the alerting mechanism can operate on machine-interpretable threshold entities instead of relying on hard-coded values, making the validation process more traceable and adaptable to future regulatory or operational requirements.

To implement the alerting logic in the current demonstrator (see Section 4.3), four threshold individuals are modelled, corresponding to the 2030 daily limit values and information and alert thresholds for PM2.5 and PM10. Annual limit values are not included in the demonstrator because they are intended for long-term compliance assessment and are therefore not deemed suitable for short-term operational alerting. Figure 3 reports the graph representation of one threshold instance, while Table 2 summarises the four threshold individuals modelled in XEMO and the corresponding values used in the demonstrator.

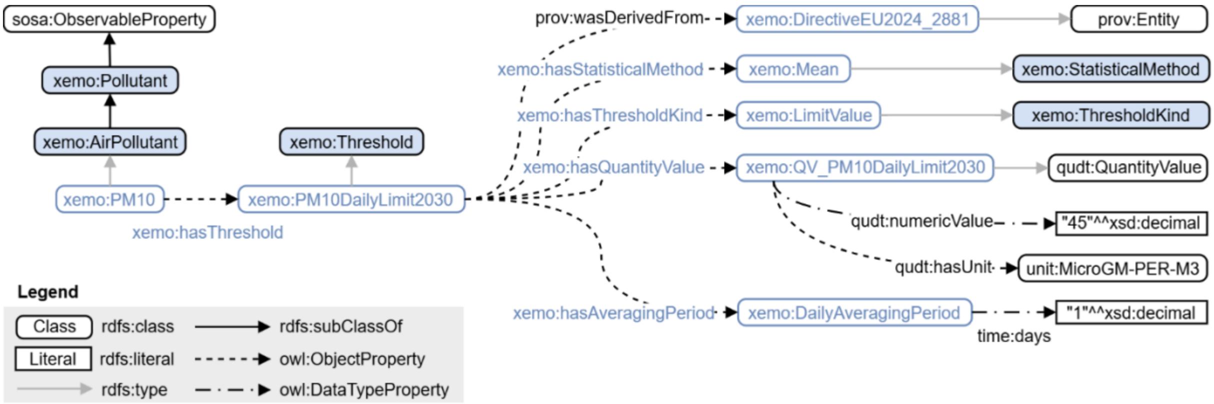

Figure 3: Graph representation of the PM10 daily mean threshold applicable from 1 January 2030 under Directive (EU) 2024/2881, modelled in XEMO (original classes, properties, and individuals are shown in blue).

Table 2: Threshold instances modelled in XEMO. The default prefix ":" denotes the XEMO namespace; ":QV_*" indicates the quantity value individual associated with each threshold instance, where "" corresponds to the threshold instance name.*

:Threshold	**:PM2_5Daily Limit2030**	**:PM2_5Alert Threshold2030**	**:PM10Daily Limit2030**	**:PM10Alert Threshold2030**
:isThresholdOfPollutant	:PM2_5		:PM10	
prov:wasDerivedFrom	:DirectiveEU2024_2881			
:hasStatisticalMethod	:Mean			
:hasThresholdKind	:LimitValue	:AlertThreshold	:LimitValue	:AlertThreshold
:hasQuantityValue	:QV_*			
qudt:numericValue	"25"^^xsd:decimal	"50"^^xsd:decimal	"45"^^xsd:decimal	"90"^^xsd:decimal
qudt:hasUnit	unit:MicroGM-PER-M3			
:hasAveragingPeriod	:DailyAveraging Period	:HourlyAveraging Period	:DailyAveraging Period	:HourlyAveraging Period
time:days / time:hours	"1"^^xsd:decimal			

4.3 Alerting logic

The modelled thresholds are used to define an operational alerting logic for a DT-oriented service based on the continuous processing of incoming PM observations. Since official regulatory assessment is based on fixed averaging periods, whereas site management requires timely updates to be effective, the demonstrator relies on updated reference values computed over moving time windows, here referred to as *rolling* values, which are then compared with the modelled thresholds. As shown in Table 3, the rolling 24-hour mean is used to evaluate daily threshold exceedance conditions, thereby remaining consistent with the required regulatory assessment based on daily average concentrations. However, since critical emission peaks may require intervention before they significantly affect the 24-hour mean, a rolling 1-hour mean is introduced as an early-warning indicator, enabling more timely mitigation actions. These rolling values therefore do not replace the official regulatory assessment but translate daily threshold values into actionable site-management triggers.

In the demonstrator, new measurements are generated, stored, processed, and assessed for exceedance conditions every five minutes. Accordingly, for each sensor, the rolling 1-hour and 24-hour means average the last 12 and 288 observations, respectively. Since the rolling 1-hour mean is used as an early-warning indicator for transient peaks, its trigger conditions are defined with higher tolerance levels than the 24-hour mean, which remains the main reference for daily threshold exceedance conditions. These demonstrative 1-hour trigger values are therefore intended to be refined or replaced once dedicated short-term criteria are made available by regulatory bodies or defined on the basis of real monitoring campaign data. The alerting state associated with the construction area in which the sensor is deployed is then assigned according to the traffic-light scheme reported in Table 3, which associates each severity level with an appropriate management response of increasing intensity. The assigned state corresponds to the highest severity condition satisfied by either the rolling 24-hour or the rolling 1-hour indicator.

Table 3: Alerting scheme adopted in the demonstrator for assigning alert states from rolling PM indicators and associating them with progressive management actions (DLT = daily limit threshold; AT = alert threshold).

Alert state	Rolling 24h condition	Rolling 1h condition	Management action
Green	< 70% DLT PM2.5 < 17.5 µg/m³ PM10 < 31.5 µg/m³	< 100 % DLT PM2.5 < 25 µg/m³ PM10 < 45 µg/m	Normal operation – Continue ordinary operations and standard mitigation measures (e.g., water spraying).
Yellow [Warning]	≥ 70% and < 100% DLT 17.5 ≤ PM2.5 < 25 µg/m³ 31.5 ≤ PM10 < 45 µg/m³	≥ 100% DLT and < AT 25 ≤ PM2.5 < 50 µg/m³ 45 ≤ PM10 < 90 µg/m³	Additional mitigation required - Notify site management and intensify mitigation measures.
Orange [Action]	≥ 100% DLT and < AT 25 ≤ PM2.5 < 50 µg/m³ 45 ≤ PM10 < 90 µg/m³	≥ AT PM2.5 ≥ 50 µg/m³ PM10 ≥ 90 µg/m³	Activity stop: Temporarily suspend the specific activities, increase mitigation measures, and record exceedance.
Red [Emergency]	≥ AT PM2.5 ≥ 50 µg/m³ PM10 ≥ 90 µg/m³	-	Emergency: Stop all dust-generating activities until alert return green; activate short-term action procedure, including external notification.

5. DEMONSTRATOR

To assess the feasibility of an alerting service based on the proposed DT framework, this study further develops the demonstrator system initially adopted to evaluate XEMO capabilities (Bruttini, Hagedorn, et al., 2026). The demonstrator is used to reproduce a synthetic construction site setting and PM management scenario, in which incoming PM observations are connected with site context, modelled thresholds, and validation-based alerting mechanisms. In this sense, it acts as a simplified operational proxy of a future DT system, focusing on the minimum components required to detect PM exceedance conditions, trace them to the relevant site entities, and support informed management actions.

Building on the preliminary XEMO evaluation pipeline, the demonstrator is extended towards a more explicit DT-oriented alerting workflow. It is used to test how synthetic PM observations can be generated, stored, semantically represented, integrated into a knowledge graph, and evaluated against the threshold logic defined in Section 4. The use of synthetic data allows this workflow to be tested independently from sensor deployment constraints, while preserving the structure required for later substitution with real monitoring streams from a construction site case study.

The scenario reproduced in the demonstrator represents a minimal construction site monitoring configuration, as shown in Figure 4. It consists of one urban roadwork site divided into three construction areas. Each area is associated with one construction activity and monitored by one PM sensor deployed within the area. For each sensor, PM10 and PM2.5 observations are generated every five minutes. Each observation is associated with the information required for semantic representation and validation, including timestamp, pollutant, sensor, construction area, related activity, feature of interest, and measured value. This setup supports the detection of threshold exceedances and their contextualisation with respect to the corresponding site condition.

The synthetic PM observations are generated to reproduce plausible environmental time series. The generator combines deterministic and stochastic components. Diurnal baseline profiles reproduce daily air-quality patterns, with lower concentrations during night-time periods and higher values during working hours. Sensor-specific scaling factors introduce spatial heterogeneity among the three monitored areas, reflecting different emission intensities associated with different activities and site zones. Work schedule effects reduce baseline concentrations during periods of lower site activity, while a sinusoidal modifier approximates smooth weather-related variation. Temporary construction activity spikes are introduced during working hours to simulate short-term emission peaks associated with dust-generating operations such as excavation, machinery movement, material transport, and roadwork activities.

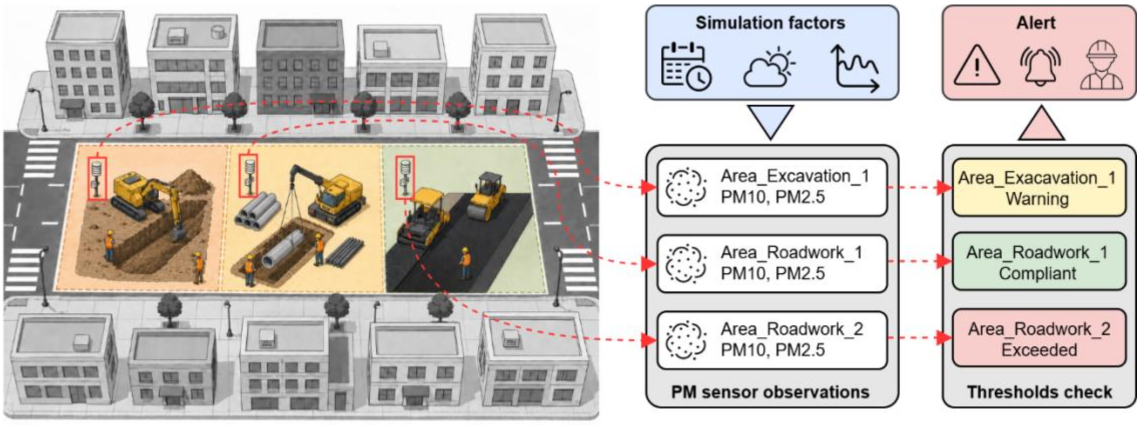

Figure 4: Synthetic urban roadwork scenario adopted in the demonstrator for PM monitoring and management.

Temporal continuity is preserved by blending each new value with the previous measurement and the current baseline condition. Long-term trend dynamics represent gradual changes in environmental conditions during the simulated period. PM2.5 values are generated in correlation with PM10 values through a variable ratio, preserving a plausible relation between the two particulate fractions. Gaussian noise and sensor-specific internal states simulate measurement variability and independent sensor behaviour, while the generated values are bounded to 5–500 µg/m³ for PM10 and 2–250 µg/m³ for PM2.5 to maintain concentrations within ranges plausible for the scenario.

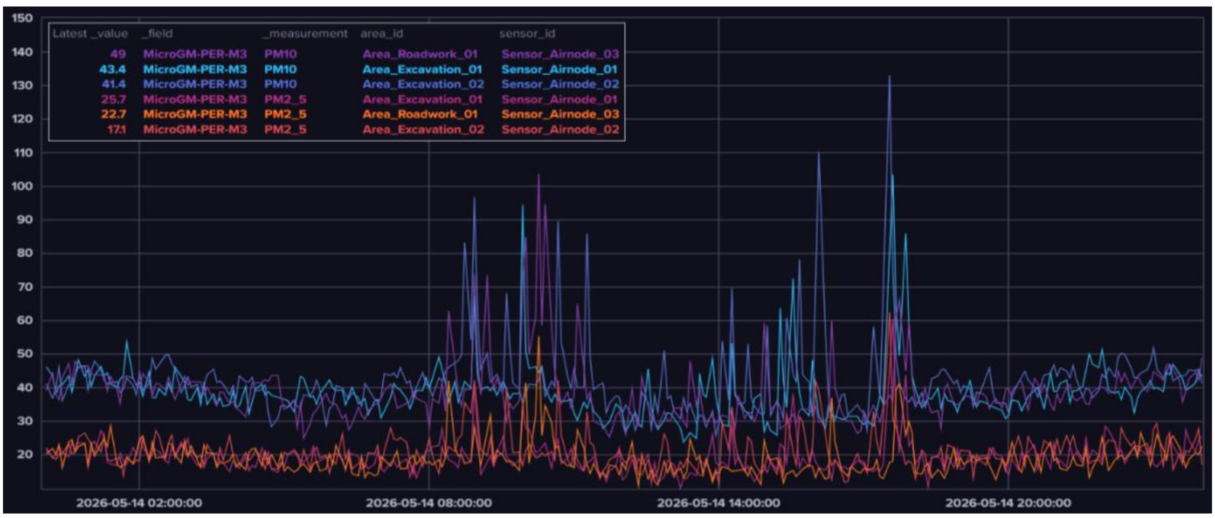

Figure 5: Synthetic PM10 and PM2.5 data stream generated for the three demonstrative sensors over a 24-hour period, from 00:00 to 24:00 on 14 May 2026 (visualised in InfluxDB).

Figure 5 shows the resulting synthetic PM data stream for the three monitored sensors over a 24-hour period, from 00:00 to 24:00 on 14 May 2026. The graph shows lower concentrations during non-working hours, increased variability and peaks during daytime construction activity, and different concentration patterns across the three monitored areas. The generated dataset provides a controlled and traceable input stream for testing the demonstrator pipeline, including time-series storage, semantic transformation, knowledge graph integration, and SHACL-based alert validation.

5.1 Architecture and implementation

Building on the synthetic scenario and PM observation stream described above, the demonstrator is implemented as a lightweight data pipeline for testing the alerting service defined in the proposed framework. As summarised in Figure 6, the workflow covers synthetic PM observation generation, time-series storage, semantic transformation of processed time-series data, integration with site and threshold information in a knowledge graph, SHACL-based validation, and alert report retrieval.

The first component of the pipeline is the Python-based *ingestor* script, which generates the synthetic PM data stream and writes it to an InfluxDB instance dedicated to time-series storage. As described in the previous subsection, the script simulates PM10 and PM2.5 observations for the sensors deployed in the three construction areas of the demonstrative roadwork scenario, producing one record for each pollutant and sensor every five minutes. Each record includes the pollutant type, unit of measure, measured value, timestamp, construction area identifier, and sensor identifier. The InfluxDB instance preserves the raw sensor-like data stream and enables the retrieval of recent measurements for each sensor and pollutant. This choice reflects the intended logic of a future DT system, where high-frequency monitoring data are stored in a dedicated time-series infrastructure, while selected observations and derived indicators are transferred to the semantic layer for contextual interpretation, validation, and decision support.

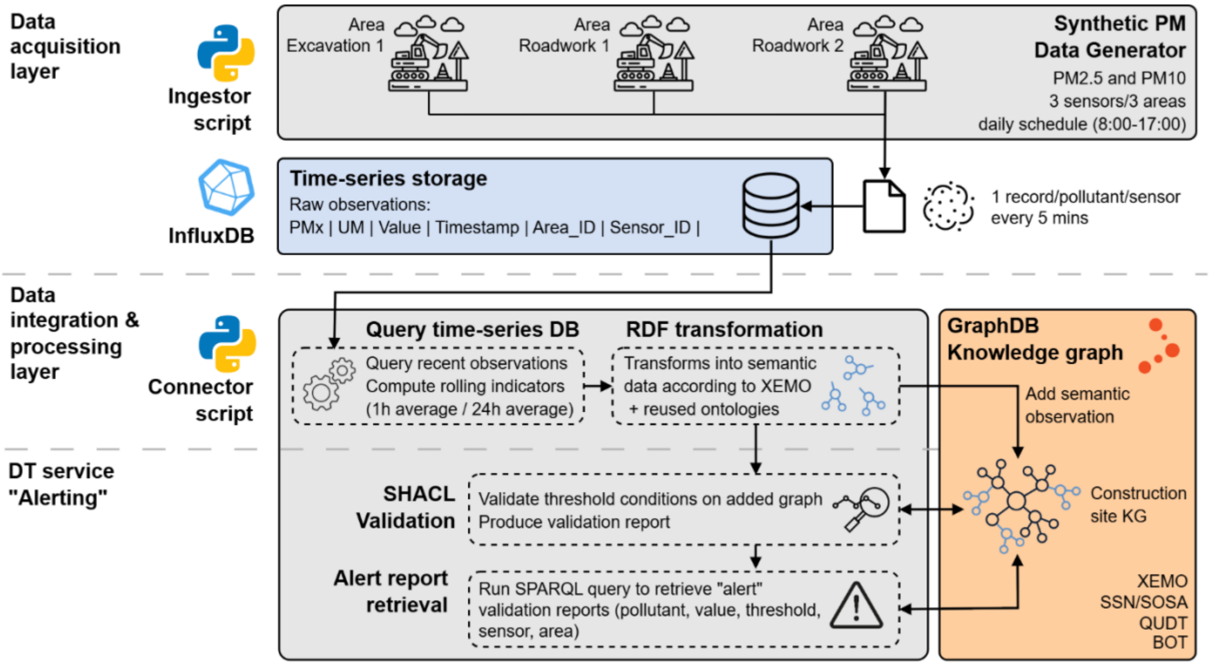

Figure 6: Demonstrator architecture and implementation workflow.

A second Python component, here referred to as the *connector* script, links data acquisition with data integration and processing. The script queries the InfluxDB instance every 60 seconds to retrieve the most recent observations and compute the rolling indicators required by the alerting logic defined in Section 4. The rolling 1-hour and 24-hour values are calculated from the most recent 12 and 288 five-minute observations, respectively. These derived values constitute the operational indicators used for threshold comparison, while the raw measurements remain stored in the time-series database. The one-minute execution interval is adopted as a polling and validation frequency, allowing the demonstrator to check for newly available observations at a higher rate than the synthetic sampling interval and to remain compatible with future monitoring configurations based on different acquisition frequencies.

The *connector* script then transforms the processed observations into RDF triples according to XEMO and the reused ontologies adopted in the semantic model. The generated triples are uploaded to a GraphDB repository, which is used as the knowledge graph management environment of the demonstrator. GraphDB was selected to store, organise, and query the RDF data required by the alerting workflow within a single integrated knowledge graph, where construction site entities, activities, sensors, pollutants, threshold entities, dynamic observations, and SHACL validation reports are represented and connected.

The graph is initially populated with the predefined entities required by the demonstrator. These include the construction site, construction areas, activities, deployed sensors, monitored pollutants, and the threshold entities described in Section 4. The entities are connected to represent the site organisation and the associations between areas and activities. The graph also identifies areas and activities as emission sources and represents the pollutants emitted and observed in the scenario, together with the threshold conditions used for alerting. This predefined structure provides the context for representing, validating, and querying the dynamic PM observations.

During pipeline execution, the graph is progressively enriched with RDF representations of the processed PM observations derived from the time-series stream. These correspond to the rolling indicators computed by the connector script and used for threshold comparison, while the original measurements remain stored in the InfluxDB instance. Each processed PM observation is represented according to the SOSA observation pattern and linked to the corresponding sensor, observed pollutant, emission source, result time, and quantitative result. The measured concentration is represented through QUDT as a quantity value with explicit numerical value and unit of measure, allowing the indicator value to be accessed during SHACL validation and SPARQL querying.

Once the RDF data have been uploaded to GraphDB, the connector script triggers SHACL validation against the shapes prepared for the alerting workflow. In the demonstrator architecture, SHACL validation is the processing step through which threshold-related alerting conditions are detected from the semantically integrated data. The validation produces SHACL validation reports, which identify the observations satisfying the modelled alert conditions and provide the basis for retrieving the related contextual information. The detailed structure of the SHACL shapes and validation reports is presented in Section 5.2.

The final component of the demonstrator consists of a SPARQL query used to retrieve the SHACL validation report and the related alert context from the knowledge graph. For each detected condition, the query returns the pollutant, measured value, threshold value, sensor, construction area, and associated emission source, completing the workflow from PM data acquisition to validation-based alert report retrieval.

5.2 SHACL-based validation

The SHACL-based validation step assesses whether the semantic workflow can support executable alerting within the proposed DT framework. It combines threshold modelling, RDF-based representation of processed PM observations, and SHACL validation to produce queryable validation reports. Accordingly, the validation step has two main objectives:

- identification of PM alert conditions from processed PM observations;
- contextual reconstruction of each detected condition to support decision-making.

The validation logic is implemented through SHACL-SPARQL constraints. Consistently with the alerting logic defined in Section 4, the validation set comprises five SHACL shapes covering the warning levels requiring management attention for 24 h rolling and hourly indicators. Each shape compares the processed PM value with the corresponding threshold condition and generates a validation result within the SHACL validation report when the condition is satisfied. The shape reported in Listing 1 targets `sosa:Observation` instances and applies the constraint only to observations associated with the 24 h rolling procedure. For each observation, the rule retrieves the observed pollutant, feature of interest, measured value, and unit of measure. It then retrieves the limit value and alert threshold associated with the same pollutant, with the orange warning condition satisfied when the measured value is equal to or higher than the limit value and lower than the alert threshold. When this condition is met, the corresponding warning is recorded in the generated SHACL validation report.

Listing 1: SHACL shape for orange warning detection based on the 24 h rolling PM indicator.

```
@prefix sh:    <http://www.w3.org/ns/shacl#> .
@prefix time:  <http://www.w3.org/2006/time#> .
@prefix xemo:  <https://example.org/xemo#> .
@prefix demo:  <https://example.org/xemo/demo#> .
@prefix sosa:  <http://www.w3.org/ns/sosa/> .
@prefix qudt:  <http://qudt.org/schema/qudt/> .
demo:OrangeWarningDailyShape
    a sh:NodeShape ;
    sh:targetClass sosa:Observation ;
    sh:severity sh:Violation ;
    sh:sparql [
        a sh:SPARQLConstraint ;
        sh:message "Orange warning: Immediate action is required, including temporary suspension of high dust-generating activities and implementation of intensified mitigation measures." ;
        sh:select """
            PREFIX time: <http://www.w3.org/2006/time#>
            PREFIX xemo: <https://example.org/xemo#>
            PREFIX sosa: <http://www.w3.org/ns/sosa/>
            PREFIX qudt: <http://qudt.org/schema/qudt/>
            SELECT $this ?feat ?pollutant ?value ?threshold ?alert ?unit
            WHERE {
                $this a sosa:Observation ;
                      sosa:observedProperty ?pollutant ;
                      sosa:hasFeatureOfInterest ?feat ;
                      sosa:hasResult/qudt:hasUnit ?unit ;
                      sosa:hasResult/qudt:numericValue ?value ;
                      sosa:hasProcedure xemo:DailyMean .
                ?t xemo:isThresholdOfPollutant ?pollutant ;
                   xemo:hasThresholdKind xemo:LimitValue ;
                   xemo:hasAveragingPeriod/time:days 1 ;
                   xemo:hasQuantityValue/qudt:numericValue ?threshold .
                ?a xemo:isThresholdOfPollutant ?pollutant ;
                   xemo:hasThresholdKind xemo:AlertThreshold ;
                   xemo:hasAveragingPeriod/time:days 1 ;
                   xemo:hasQuantityValue/qudt:numericValue ?alert .
                FILTER(?value >= ?threshold && ?value < ?alert)
            }
        """ ;
    ] .
```

The same structure is adapted for the SHACL shapes related to the other warning levels and averaging procedures reported in Appendix A. For the 24 h rolling indicator, the yellow warning identifies values between 70% of the limit value and the limit value, while the red warning identifies values equal to or higher than the alert threshold. For the hourly indicator, the shapes target observations associated with the hourly mean procedure and apply the short-term warning logic defined in Section 4.

The described validation process produces reports that are integrated into the graph and can subsequently be queried. In the demonstrator, a SPARQL query accesses the generated report and reconstructs the corresponding alert context. Starting from the observation identified in the report, the query retrieves the associated pollutant, timestamp, threshold value, measured value, unit of measure, feature of interest, and warning message from the validation report. Figure 7 shows an example of this retrieval step in GraphDB.

The retrieved information provides the semantic basis for alert-oriented services. At demonstrator level, it supports the production of alert reports containing the contextual information required to support decision-making. In a future DT implementation, the same mechanism could enable dashboard alerts, worker or site manager notifications, compliance logs, diagnosis workflows, and the activation of automated mitigation measures by connecting SHACL-based detection with SPARQL-based context retrieval within a single graph-based workflow.

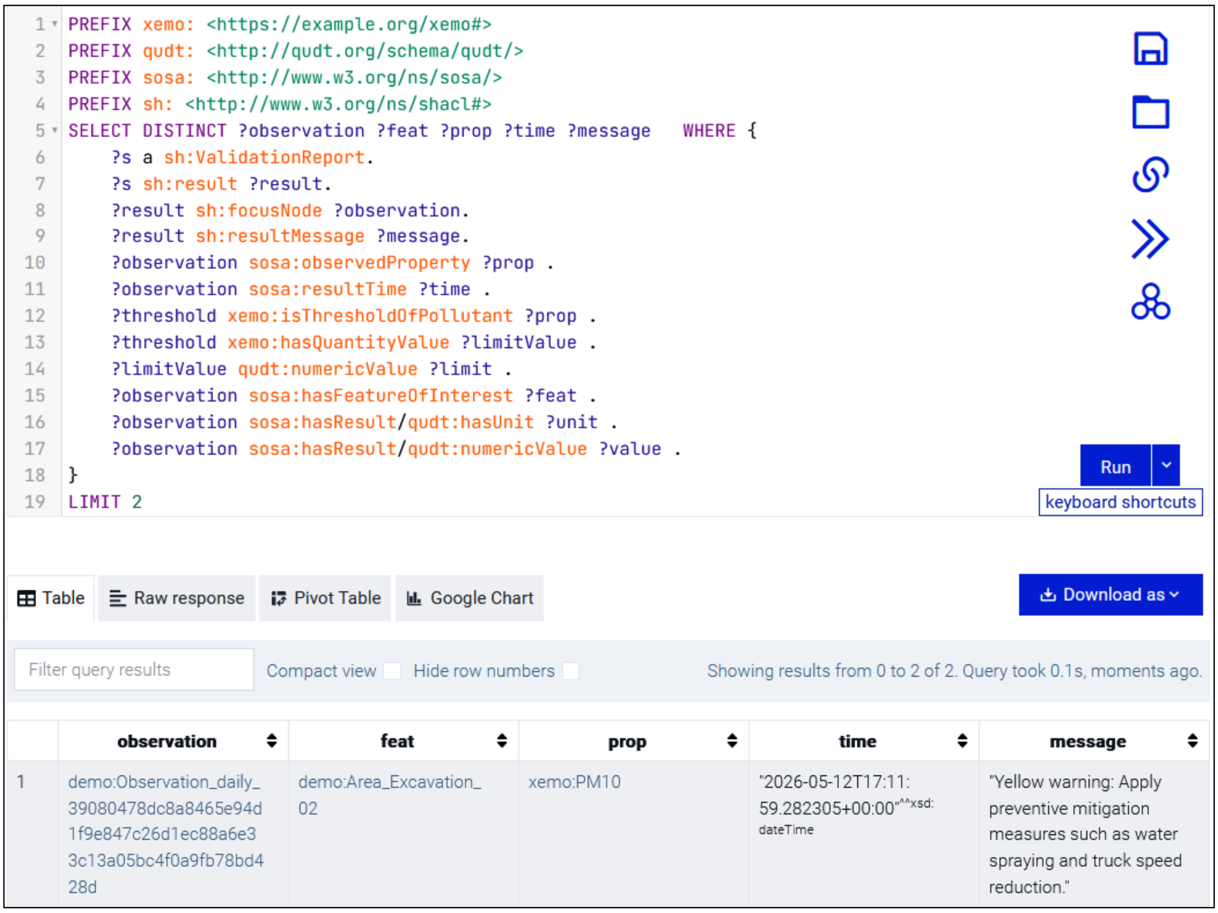

Figure 7: Retrieval of SHACL validation results in GraphDB for PM warning detection and alert-context reconstruction.

6. DISCUSSION

The demonstrator represents a first implementation-oriented step towards the semantic DT direction outlined in the proposed framework for construction site PM emission management. Starting from XEMO as the semantic backbone, this work extends the previous demonstrator from an ontology capability assessment towards the evaluation of an alerting service. This addresses the need for more structured DT workflows able to link processed PM observations to threshold entities within a semantic structure, evaluate warning conditions through SHACL, and trace the resulting alerts back to the relevant construction areas and emission sources.

A central aspect of this extension is the semantic modelling of PM threshold entities. The background analysis showed that PM threshold logic for construction site alerting is difficult to formalise, especially because available reference values mainly derive from environmental air quality regulations and do not directly capture short term, task-specific, and location-dependent worker exposure. The proposed threshold model does not address this gap and does not define occupational exposure limits, which remain the responsibility of competent regulatory bodies. It shows, however, that threshold entities can be explicitly represented, linked to pollutants and averaging procedures, and used as actionable elements within SHACL-based validation rules. This makes threshold logic computationally available for alerting services and keeps its assumptions explicit, traceable, and open to transparent revision according to different management requirements or future regulatory values.

The implemented workflow also shows the feasibility of SHACL-based semantic alerting at prototype level. Processed PM indicators are transformed into RDF, integrated with the site and threshold graph, and validated against warning conditions. The SHACL shapes themselves can also be represented and managed as RDF resources in a dedicated shapes graph, favouring reuse, versioning, and transfer across different projects or monitoring scenarios. The resulting validation reports can then be stored, queried, and used to reconstruct the corresponding alert context through SPARQL. Warning conditions can therefore be detected and traced back to the monitored pollutant, measured concentration, threshold condition, sensor, construction area, and associated emission source. This traceability connects numerical PM conditions with the site entities needed to support interpretation and management action.

The current semantic infrastructure remains limited but extensible. XEMO is still at an early stage of development, and the demonstrator instantiates only a limited set of entities and relations. Its graph-based structure provides a basis for progressively linking additional heterogeneous data, including weather conditions, construction activities, mitigation measures, exposed receptors, and other contextual information represented through existing or future ontologies. This extensibility is relevant for moving from basic alert detection towards more advanced reasoning and decision-support use cases within a semantic DT for construction site PM emission management.

The implemented pipeline also clarifies the current coverage of the proposed DT framework. It covers data collection, storage, integration, linking, processing, semantic validation, and alert-context retrieval. It also provides a basis for data visualisation through GraphDB query outputs. Higher-level DT capabilities remain outside the current implementation, including learning from historical patterns, predicting future PM conditions, recommending mitigation actions, and supporting automated or semi-automated intervention workflows.

6.1 Limitations and future work

Among the main limitations of this work, the first concerns the use of synthetic PM observations generated from predefined assumptions instead of real sensor data streams. This choice allows the alerting workflow to be tested under controlled conditions, but it does not validate the system against the variability, noise, calibration issues, and data gaps that characterise real construction site monitoring. Second, the demonstrative construction site is represented through a simplified roadwork scenario composed of three construction areas, fixed sensors, and predefined emission sources. This scenario is sufficient to test traceability between alerts and site entities, but it does not represent the spatial, temporal, and organisational complexity of real construction sites, where construction areas, activities, emission sources, and sensors may be more numerous and vary over time and across project phases, even in relatively small projects. Third, the validation logic is limited to PM10 and PM2.5 and to simplified threshold conditions defined for the demonstrator. Although rolling 1 h and 24 h indicators are computed before RDF transformation, the current implementation does not yet address more advanced uncertainty handling, sensor reliability assessment, or adaptive threshold logic, which may require more complex processing methods and additional data quality-control procedures. The warning conditions are used to assess the feasibility of the semantic and computational workflow, without constituting a complete occupational or environmental compliance model.

A further limitation concerns the technological maturity and scalability of the implemented infrastructure. The demonstrator remains a prototype and cannot be considered a production-ready semantic DT system. It does not test long-term data accumulation, high-frequency multi-sensor streams, larger graph sizes, query performance under increasing data volume, or deployment requirements for real-time operation. These aspects are critical because a real DT infrastructure would need to define which data remain in time-series storage, which processed indicators are transferred to the graph, how validation reports are retained and managed, and how query performance is preserved as the system grows. Assessing computational scalability and operational robustness will therefore require tests with larger data volumes and more demanding operational workloads.

Future work should extend the demonstrator towards real monitoring and deployment conditions. This direction is already being pursued within the broader research effort in which this contribution is situated, through the preparation and evaluation of a real case study based on sensor-based PM data from an urban construction site located in Italy. Further work should include the integration of live sensor streams, calibration and quality-control procedures, weather and activity data, more detailed representations of exposed receptors and mitigation actions, and dashboard interfaces for site managers and workers. It should also investigate data storage and handling strategies for scalable semantic DT operation, including retention policies, graph update mechanisms, validation report management, and performance-oriented query design. Finally, future extensions should address predictive modules and recommendation mechanisms, allowing the framework to move from alert detection towards prediction and decision-support capabilities for construction site PM emission management.

7. CONCLUSIONS

Construction site PM emission management requires approaches able to connect monitoring data, site context, threshold information, and decision-support services. This need is particularly relevant in urban contexts, where PM emissions affect workers, surrounding communities, and the liveability of dense built environments, with implications for health and well-being, sustainable infrastructure, and sustainable cities. In this perspective, the paper contributes to a broader research direction aligned with UN SDG 3, SDG 9, and SDG 11 (United Nations, 2015a, 2015b, 2015c) by addressing how semantic technologies can support PM alerting towards a Digital Twin for urban construction site emission management.

Building on a previous high-level DT framework for PM emission management on urban construction sites (Bruttini, Sorbi, et al., 2026a), this paper focused on the semantic and validation mechanisms required to implement part of that framework. In a recent related contribution within the same research direction, the authors developed XEMO (Bruttini, Hagedorn, et al., 2026) to provide the semantic backbone for describing the construction site environment targeted by PM emission management, including construction site entities, pollutants, sensors, observations, emission sources, and thresholds. The present extension further develops the demonstrator previously implemented for the XEMO capability assessment towards an explicit alerting workflow. In this workflow, processed PM observations are transformed into RDF, integrated into a knowledge graph, and evaluated through SHACL validation according to the threshold-based warning logic proposed in this paper.

The main contribution of the paper lies in the modelling and operational use of threshold entities within a semantic alerting workflow for PM emission management. The work showed how threshold information can be represented as explicit semantic entities, linked to pollutants, averaging procedures, quantitative values, units, and regulatory sources, and used within SHACL-based validation rules. The demonstrator confirmed the prototype-level feasibility of this approach by generating synthetic PM10 and PM2.5 observations, storing them as time-series data, transforming processed indicators into semantic representations, integrating them with site and threshold information, and validating them against warning conditions. The resulting validation reports could then be queried to retrieve the alert context, including the pollutant, measured concentration, threshold condition, sensor, construction area, and associated emission source. This makes threshold logic available for computational processing and shows how PM observations can be transformed into semantically interpretable data to support alert reporting and management action.

This paper provides an intermediate but necessary step towards a semantic DT for construction site PM emission management. The next development stage will address the current limitations through deployment with real construction site monitoring data, further assessment of the technological infrastructure under real data volumes and operating conditions, and extension of the semantic model to richer contextual information. On this basis, the work establishes a reusable foundation for alerting and more advanced decision-support services.

DATA AVAILABILITY

The demonstrator uses synthetic PM observations generated at runtime. The scripts used to generate these observations, populate the semantic knowledge graph, and execute the recurrent SHACL validation workflow, together with the SHACL shapes, are openly available at: https://github.com/AlessandroBruttini/XEMO-PM-SHACL-Alerting-demonstrator. The XEMO ontology used by the demonstrator is maintained in a separate public repository: https://github.com/AlessandroBruttini/XEMO.

ACKNOWLEDGEMENTS

This paper is an extended version of our previous work, which was presented at the Smart and Sustainable Built Environment Conference, Lille – France | 3rd to 5th November 2025. The authors acknowledge the support and feedback from the conference chairs, Prof. Farzad Rahimian, Professor Abiola Akanmu and Dr M. Reza Hosseini, along with their team, throughout the peer review process of SASBE 2025 and during the conference, which helped improve our submissions. This research was supported by the project *Digital Twin development for the resilient* *management of minor historical centres (DiTHiCe)*, funded under the Italian national programme Projects of Relevant National Interest, CUP: E53C24002740006. No human or animal participants were involved in this study.

REFERENCES

Adreani, L., Bellini, P., Fanfani, M., Nesi, P., & Pantaleo, G. (2024). Smart City Digital Twin Framework for Real-Time Multi-Data Integration and Wide Public Distribution. *IEEE Access*, *12*, 76277–76303. https://doi.org/10.1109/ACCESS.2024.3406795

Azarmi, F., Kumar, P., Marsh, D., & Fuller, G. (2016). Assessment of the long-term impacts of PM_{10} and $PM_{2.5}$ particles from construction works on surrounding areas. *Environmental Science: Processes & Impacts*, *18*(2), 208–221. https://doi.org/10.1039/C5EM00549C

Berners-Lee, T. (2006). *Linked Data—Design Issues*. https://www.w3.org/DesignIssues/LinkedData.html

Boje, C., Kubicki, S., Guerriero, A., Rezgui, Y., & Zarli, A. (2022). Digital twins for the built environment. In *Buildings and Semantics* (pp. 179–199). CRC Press. DOI: 10.1201/9781003204381-10

Borrmann, A., König, M., Koch, C., & Beetz, J. (2018). Building Information Modeling: Why? What? How? In A. Borrmann, M. König, C. Koch, & J. Beetz (Eds), *Building Information Modeling* (pp. 1–24). Springer International Publishing. https://doi.org/10.1007/978-3-319-92862-3_1

Brostrøm, A., Thalmann, J., Liisberg, J. B., Husovská, F., Møller, S. H., Rasmussen, J. T., Jensen, T. N., Jensen, S. B., Jensen, K. A., Cole-Hunter, T., & Fonseca, A. S. (2025). Task specific assessment of particle exposure and low-cost sensor performance in indoor construction environments. *Atmospheric Environment: X*, *26*, 100336. https://doi.org/10.1016/j.aeaoa.2025.100336

Bruttini, A., Hagedorn, P., Getuli, V., Capone, P., & König, M. (2026). XEMO: Construction Site Emission Management Ontology for Particulate Matter. *Proceedings of the 33rd EG-ICE International Workshop on Intelligent Computing in Engineering, 1st-3rd July, Padua, Italy (In Press)*.

Bruttini, A., Sorbi, T., Getuli, V., & Capone, P. (2026a). A Digital Twin Framework for Monitoring and Managing PM Emissions on Urban Construction Sites. In F. Rahimian, M. R. Hosseini, A. Akanmu, Z. Lafhaj, & L. Ducoulombier (Eds), *Proceedings of the International Conference on Smart and Sustainable Built Environment* (Vol. 844, pp. 87–100). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-8489-5_8

Bruttini, A., Sorbi, T., Getuli, V., & Capone, P. (2026b). Towards Digital Twins for PM Emission Management on Urban Construction Sites: A Systematic Review. In F. Rahimian, M. R. Hosseini, A. Akanmu, Z. Lafhaj, & L. Ducoulombier (Eds), *Proceedings of the International Conference on Smart and Sustainable Built Environment* (Vol. 844, pp. 101–116). Springer Nature Singapore. https://doi.org/10.1007/978-981-95-8489-5_9

Cheriyan, D., & Choi, J. (2020). A review of research on particulate matter pollution in the construction industry. *Journal of Cleaner Production*, *254*, 120077. https://doi.org/10.1016/j.jclepro.2020.120077

Cho, S., Khan, M., Pyeon, J., & Park, C. (2021). Blockchain-Based Network Concept Model for Reliable and Accessible Fine Dust Management System at Construction Sites. *Applied Sciences*, *11*(18), 8686. https://doi.org/10.3390/app11188686

Corneli, A., & Rotilio, M. (2023). URBAN CENTERS MANAGEMENT: A DIGITAL TWIN APPROACH. *Managing the Digital Transformation of Construction Industry*. CONVR 2023 - 23rd International Conference on Construction Applications of Virtual Reality Managing the Digital Transformation of Construction Industry. https://doi.org/10.36253/979-12-215-0289-3.122

Cox, S., & Little, C. (2022). *Time Ontology in OWL*. https://www.w3.org/TR/owl-time/

European Parliament and Council. (2008). *Directive 2008/50/EC of the European Parliament and of the Council of 21 May 2008 on ambient air quality and cleaner air for Europe*. https://eur-lex.europa.eu/eli/dir/2008/50/oj

European Parliament and Council. (2024). *Directive (EU) 2024/2881 of the European Parliament and of the Council of 23 October 2024 on ambient air quality and cleaner air for Europe*. https://eur-lex.europa.eu/eli/dir/2024/2881/oj

FAIRsharing Team. (2015). *FAIRsharing record for: Quantities, Units, Dimensions and Types* [Data set]. FAIRsharing. https://doi.org/10.25504/FAIRSHARING.D3PQW7

Fang, X., Chang, R., Zuo, J., Zhang, W. E., Zou, Y., & Li, K. (2025). How do environmental and operational factors impact particulate matter dynamics in building construction? - Insights from real-time sensing. *Journal of Environmental Management*, *380*, 125098. https://doi.org/10.1016/j.jenvman.2025.125098

Farghaly, K., Soman, R. K., & Zhou, S. A. (2023). The evolution of ontology in AEC: A two-decade synthesis, application domains, and future directions. *Journal of Industrial Information Integration*, *36*, 100519. https://doi.org/10.1016/j.jii.2023.100519

Gill, E. Z., Cangelmi, L., Cellini, P., Zaldei, A., Draganov, I. R., Cardone, D., & Amelio, A. (2025). *Forecasting Air Pollution at Construction Sites: An Extended Comparison of Machine Learning Approaches*. https://doi.org/https://doi.org/10.18485/mtts_mr.2025.31.2.5

Grieves, M. W. (2005). Product lifecycle management: The new paradigm for enterprises. *International Journal of Product Development*, *2*(1/2), 71. https://doi.org/10.1504/IJPD.2005.006669

Guo, X., Wang, Y., Mei, S., Shi, C., Liu, Y., Pan, L., Li, K., Zhang, B., Wang, J., Zhong, Z., & Dong, M. (2022). Monitoring and modelling of PM2.5 concentration at subway station construction based on IoT and LSTM algorithm optimization. *Journal of Cleaner Production*, *360*, 132179. https://doi.org/10.1016/j.jclepro.2022.132179

Hagedorn, P., Petrova, E., & König, M. (2025). A System-of-Systems Approach for Deploying Containerized Construction Digital Twins Using Linked Data. In A. Francis, E. Miresco, & S. Melhado (Eds), *Advances in Information Technology in Civil and Building Engineering. ICCCBE 2024.* (Vol. 628, pp. 478–491). Springer Nature Switzerland. https://doi.org/10.1007/978-3-031-84208-5_36

Haller, A., Janowicz, K., Cox, S., Le Phuoc, D., Taylor, K., Lefrancois, M., Atkinson, R., Garcia-Castro, R., Lieberman, J., & Stadler, C. (2017). *Semantic Sensor Network Ontology*. https://www.w3.org/TR/vocab-ssn/

Ham, Y.-B., Cheriyan, D., Kim, H.-U., Han, J.-G., Kim, Y. H., Janani Priyanka, P. R., & Choi, J. (2024). Particulate matter reduction efficiency analysis of sprinkler system as targeted control measures for construction activity. *Heliyon*, *10*(7), e27765. https://doi.org/10.1016/j.heliyon.2024.e27765

Han, M., Zhang, J., Huang, Y., Xu, J., Chen, X., & Chen, B. M. (2025). Enhancing worker monitoring and management on large-scale construction sites with UAVs and digital twin modeling. *Automation in Construction*, *174*, 106108. https://doi.org/10.1016/j.autcon.2025.106108

Hong, J., Kang, H., Jung, S., Sung, S., Hong, T., Park, H. S., & Lee, D.-E. (2020). An empirical analysis of environmental pollutants on building construction sites for determining the real-time monitoring indices. *Building and Environment*, *170*, 106636. https://doi.org/10.1016/j.buildenv.2019.106636

Jaafar, W., Xu, J., Farrar, E., Jeong, C.-H., Ganji, A., Evans, G., & Hatzopoulou, M. (2024). Challenges and opportunities of low-cost sensors in capturing the impacts of construction activities on neighborhood air quality. *Building and Environment*, *254*, 111363. https://doi.org/10.1016/j.buildenv.2024.111363

Jiang, L., Shi, J., & Wang, C. (2022). Multi-ontology fusion and rule development to facilitate automated code compliance checking using BIM and rule-based reasoning. *Advanced Engineering Informatics*, *51*, 101449. https://doi.org/10.1016/j.aei.2021.101449

Kang, H., Sung, S., Hong, J., Jung, S., Hong, T., Park, H. S., & Lee, D.-E. (2021). Development of a real-time automated monitoring system for managing the hazardous environmental pollutants at the construction site. *Journal of Hazardous Materials*, *402*, 123483. https://doi.org/10.1016/j.jhazmat.2020.123483

Khan, M., Khan, N., Skibniewski, M. J., & Park, C. (2021). Environmental Particulate Matter (PM) Exposure Assessment of Construction Activities Using Low-Cost PM Sensor and Latin Hypercubic Technique. *Sustainability*, *13*(14), 7797. https://doi.org/10.3390/su13147797

Kim, H., Kim, J., & Wang, S. (2025). Automated human-centric construction dust alert system for construction workers: Field test operation and technical review. *Journal of Computational Design and Engineering*, *12*(11), 37–52. https://doi.org/10.1093/jcde/qwaf101

Kim, H., Tae, S., Zheng, P., Kang, G., & Lee, H. (2021). Development of IoT-Based Particulate Matter Monitoring System for Construction Sites. *International Journal of Environmental Research and Public Health*, *18*(21), 11510. https://doi.org/10.3390/ijerph182111510

Kosse, S., Hagedorn, P., & König, M. (2025). Semantic Digital Twins in Construction: Developing a modular System Reference Architecture based on Information Containers. *Advanced Engineering Informatics*, *67*, 103483. https://doi.org/10.1016/j.aei.2025.103483

Lebo, T., Sahoo, S., McGuinness, D. L., Belhajjame, K., Cheney, J., Corsar, Garijo, D., Soiland-Reyes, S., Zednik, S., & Zhao, J. (2013). *PROV-O: The PROV Ontology*. https://www.w3.org/TR/prov-o/

Li, C. Z., Zhao, Y., & Xu, X. (2019). Investigation of dust exposure and control practices in the construction industry: Implications for cleaner production. *Journal of Cleaner Production*, *227*, 810–824. https://doi.org/10.1016/j.jclepro.2019.04.174

Li, X., Yang, D., Yuan, J., Donkers, A., & Liu, X. (2022). BIM-enabled semantic web for automated safety checks in subway construction. *Automation in Construction*, *141*, 104454. https://doi.org/10.1016/j.autcon.2022.104454

Lu, Y., Li, Q., Zhou, Z., & Deng, Y. (2015). Ontology-based knowledge modeling for automated construction safety checking. *Safety Science*, *79*, 11–18. https://doi.org/10.1016/j.ssci.2015.05.008

Lu, Y., Song, G., Li, P., & Wang, N. (2024). Development of an ontology for construction carbon emission tracking and evaluation. *Journal of Cleaner Production*, *443*, 141170. https://doi.org/10.1016/j.jclepro.2024.141170

Luo, Q., Huang, L., Liu, Y., Xue, X., Zhou, F., & Hua, J. (2021). Monitoring Study on Dust Dispersion Properties during Earthwork Construction. *Sustainability*, *13*(15), 8451. https://doi.org/10.3390/su13158451

Milivojević, L., Mrazovac Kurilić, S., Božilović, Z., Koprivica, S., & Krčadinac, O. (2023). Study of Particular Air Quality and Meteorological Parameters at a Construction Site. *Atmosphere*, *14*(8), 1267. https://doi.org/10.3390/atmos14081267

Milivojevic, L., Mrazovac-Kurilic, S., Sunjevic, M., Bozilovic, Z., Golubovic-Matic, D., & Cekic, Z. (2024). Modeling of particulate matter concentrations on a construction site based on IoT monitoring system. *Thermal Science*, *28*(5 Part B), 4249–4266. https://doi.org/10.2298/TSCI231209153M

Paolucci, R., Rotilio, M., Ricci, S., Pelliccione, A., & Ferri, G. (2022). A Sensor-Based System for Dust Containment in the Construction Site. *Energies*, *15*(19), 7272. https://doi.org/10.3390/en15197272

Patel, D. J., Patel, D. A., & Patel, S. (2024). Assessment of the Environmental Pollutants of Demolition Sites for Developing Real-Time Monitoring Indexes: An Empirical Analysis. *Journal of Legal Affairs and Dispute Resolution in Engineering and Construction*, *16*(1), 04523052. https://doi.org/10.1061/JLADAH.LADR-1053

Pham, T.-A., Ioannou, P. G., & Likhitruangsilp, V. (2026). Designing a BIM-enhanced digital twin architecture for construction pollution management in building renovation projects. *International Journal of Construction Management*, 1–26. https://doi.org/10.1080/15623599.2026.2651946

Pham, T.-A., Likhitruangsilp, V., Nguyen, T. A., & Ioannou, P. (2025). *Common data environment of a BIM-based digital twin for real-time construction pollution management*. https://doi.org/10.1108/ECAM-03-2025-0527

Rasmussen, M. H., Pauwels, P., Lefrancois, M., & Schneider, G. F. (2021). *Building Topology Ontology (BOT)*. https://w3c-lbd-cg.github.io/bot/

Sacks, R., Brilakis, I., Pikas, E., Xie, H. S., & Girolami, M. (2020). Construction with digital twin information systems. *Data-Centric Engineering*, *1*, e14. https://doi.org/10.1017/dce.2020.16

Schlenger, J., Pluta, K., Mathew, A., Yeung, T., Sacks, R., & Borrmann, A. (2025). Reference architecture and ontology framework for digital twin construction. *Automation in Construction*, *174*, 106111. https://doi.org/10.1016/j.autcon.2025.106111

Tao, F., Sui, F., Liu, A., Qi, Q., Zhang, M., Song, B., Guo, Z., Lu, S. C.-Y., & Nee, A. Y. C. (2019). Digital twin-driven product design framework. *International Journal of Production Research*, *57*(12), 3935–3953. https://doi.org/10.1080/00207543.2018.1443229

Tao, G., Feng, J., Feng, H., Feng, H., & Zhang, K. (2022). Reducing Construction Dust Pollution by Planning Construction Site Layout. *Buildings*, *12*(5), 531. https://doi.org/10.3390/buildings12050531

United Nations. (2015a). *Goal 03: Ensure healthy lives and promote well-being for all at all ages – Targets and Indicators. United Nations Sustainable Development Goals.* https://sdgs.un.org/goals/goal3#targets_and_indicators

United Nations. (2015b). *Goal 09: Build resilient infrastructure, promote inclusive and sustainable industrialization and foster innovation – Targets and Indicators. United Nations Sustainable Development Goals.* https://sdgs.un.org/goals/goal9#targets_and_indicators

United Nations. (2015c). *Goal 11: Make Cities and Human Settlements Inclusive, Safe, Resilient and Sustainable – Targets and Indicators. United Nations Sustainable Development Goals.* https://sdgs.un.org/goals/goal11#targets_and_indicators

W3C Linked Building Data Community Group. (2022). https://www.w3.org/community/lbd/

Wang, Y., Su, Y., Zheng, Z., Zhou, Z., & Wang, X. (2026). Bayesian-optimized CNN-LSTM neural network for predicting road construction dust concentrations. *Developments in the Built Environment*, *25*, 100843. https://doi.org/10.1016/j.dibe.2026.100843

Weil, C., Bibri, S. E., Longchamp, R., Golay, F., & Alahi, A. (2023). Urban Digital Twin Challenges: A Systematic Review and Perspectives for Sustainable Smart Cities. *Sustainable Cities and Society*, *99*, 104862. https://doi.org/10.1016/j.scs.2023.104862

Xu, Y., Wang, S., Chen, J., Liu, K., Wei, Y., Zhang, Y., Wang, J., & Jin, L. (2025). Rapid dust concentration prediction using sparse sensors during tunnel construction. *Building Simulation*, *18*(6), 1543–1559. https://doi.org/10.1007/s12273-025-1284-0

Xu, Z., Huo, H., & Pang, S. (2022). Identification of Environmental Pollutants in Construction Site Monitoring Using Association Rule Mining and Ontology-Based Reasoning. *Buildings*, *12*(12), 2111. https://doi.org/10.3390/buildings12122111

Yoon, S., Won, D., & Chi, S. (2023). Empirical studies on emission factors for real-time particulate matter 2.5 monitoring at construction sites. *Journal of Cleaner Production*, *384*, 135546. https://doi.org/10.1016/j.jclepro.2022.135546

Zhang, R., Sapkota, J., & Krall, J. (2026). Multimodal Data Fusion and Deep Learning for Predicting Air Pollution at Construction Sites. *Journal of Computing in Civil Engineering*, *40*(2), 04025132. https://doi.org/10.1061/JCCEE5.CPENG-7003

Zheng, L., Feng, Z., Liu, J., Han, L., & Wang, H. (2024). Cross-concentration calibration of low-cost sensors for effective dust monitoring at construction sites. *Journal of Aerosol Science*, *182*, 106456. https://doi.org/10.1016/j.jaerosci.2024.106456

Zhong, B., Gan, C., Luo, H., & Xing, X. (2018). Ontology-based framework for building environmental monitoring and compliance checking under BIM environment. *Building and Environment*, *141*, 127–142. https://doi.org/10.1016/j.buildenv.2018.05.046

Zhong, B., Guo, J., Zhang, L., Wu, H., Li, H., & Wang, Y. (2022). A blockchain-based framework for on-site construction environmental monitoring: Proof of concept. *Building and Environment*, *217*, 109064. https://doi.org/10.1016/j.buildenv.2022.109064

APPENDIX A: SHACL SHAPES FOR PM WARNING DETECTION

This appendix reports the additional SHACL shapes used in the demonstrator for PM warning detection. The orange warning shape for the 24 h rolling indicator is reported in Section 5.2 as the main example. The shapes below define the remaining relevant warning conditions for daily and hourly indicators.

Listing A.1: SHACL shape for yellow warning detection based on the 24 h rolling PM indicator.

```
@prefix sh:    <http://www.w3.org/ns/shacl#> .
@prefix time:  <http://www.w3.org/2006/time#> .
@prefix xemo:  <https://example.org/xemo#> .
@prefix demo:  <https://example.org/xemo/demo#> .
@prefix sosa:  <http://www.w3.org/ns/sosa/> .
@prefix qudt:  <http://qudt.org/schema/qudt/> .

demo:YellowWarningDailyShape
    a sh:NodeShape ;
    sh:targetClass sosa:Observation ;
    sh:severity sh:Warning ;
    sh:sparql [
        a sh:SPARQLConstraint ;
        sh:message "Yellow warning: Apply preventive mitigation measures such as water spraying and truck
speed reduction." ;
        sh:select """
            PREFIX time: <http://www.w3.org/2006/time#>
            PREFIX xemo: <https://example.org/xemo#>
            PREFIX sosa: <http://www.w3.org/ns/sosa/>
            PREFIX qudt: <http://qudt.org/schema/qudt/>

            SELECT $this ?feat ?pollutant ?value ?threshold ?unit
            WHERE {
                $this a sosa:Observation ;
                      sosa:observedProperty ?pollutant ;
                      sosa:hasFeatureOfInterest ?feat ;
                      sosa:hasResult/qudt:hasUnit ?unit ;
                      sosa:hasResult/qudt:numericValue ?value ;
                      sosa:hasProcedure xemo:DailyMean .

                ?t xemo:isThresholdOfPollutant ?pollutant ;
                   xemo:hasThresholdKind xemo:LimitValue ;
                   xemo:hasAveragingPeriod/time:days 1 ;
                   xemo:hasQuantityValue/qudt:numericValue ?threshold .

                FILTER(?value >= 0.7 * ?threshold && ?value < ?threshold)
            }
        """ ;
    ] .
```

Listing A.2: SHACL shape for red warning detection based on the 24 h rolling PM indicator.

```
@prefix sh:    <http://www.w3.org/ns/shacl#> .
@prefix time:  <http://www.w3.org/2006/time#> .
@prefix xemo:  <https://example.org/xemo#> .
@prefix demo:  <https://example.org/xemo/demo#> .
@prefix sosa:  <http://www.w3.org/ns/sosa/> .
```

```
@prefix qudt:  <http://qudt.org/schema/qudt/> .

demo:RedWarningDailyShape
    a sh:NodeShape ;
    sh:targetClass sosa:Observation ;
    sh:severity sh:Violation ;
    sh:sparql [
        a sh:SPARQLConstraint ;
        sh:message "Red warning: Stop all dust-generating activities until conditions return below the
alert threshold." ;
        sh:select """
            PREFIX time: <http://www.w3.org/2006/time#>
            PREFIX xemo: <https://example.org/xemo#>
            PREFIX sosa: <http://www.w3.org/ns/sosa/>
            PREFIX qudt: <http://qudt.org/schema/qudt/>

            SELECT $this ?feat ?pollutant ?value ?alert ?unit
            WHERE {
                $this a sosa:Observation ;
                      sosa:observedProperty ?pollutant ;
                      sosa:hasFeatureOfInterest ?feat ;
                      sosa:hasResult/qudt:hasUnit ?unit ;
                      sosa:hasResult/qudt:numericValue ?value ;
                      sosa:hasProcedure xemo:DailyMean .

                ?a xemo:isThresholdOfPollutant ?pollutant ;
                   xemo:hasThresholdKind xemo:AlertThreshold ;
                   xemo:hasAveragingPeriod/time:days 1 ;
                   xemo:hasQuantityValue/qudt:numericValue ?alert .

                FILTER(?value >= ?alert)
            }
        """ ;
    ] .
```

Listing A.3: SHACL shape for yellow warning detection based on the hourly PM indicator.

```
@prefix sh:    <http://www.w3.org/ns/shacl#> .
@prefix time:  <http://www.w3.org/2006/time#> .
@prefix xemo:  <https://example.org/xemo#> .
@prefix demo:  <https://example.org/xemo/demo#> .
@prefix sosa:  <http://www.w3.org/ns/sosa/> .
@prefix qudt:  <http://qudt.org/schema/qudt/> .

demo:YellowWarningHourlyShape
    a sh:NodeShape ;
    sh:targetClass sosa:Observation ;
    sh:severity sh:Warning ;
    sh:sparql [
        a sh:SPARQLConstraint ;
        sh:message "Yellow warning: Apply preventive mitigation measures such as water spraying and truck
speed reduction." ;
        sh:select """
            PREFIX time: <http://www.w3.org/2006/time#>
```

```
            PREFIX xemo: <https://example.org/xemo#>
            PREFIX sosa: <http://www.w3.org/ns/sosa/>
            PREFIX qudt: <http://qudt.org/schema/qudt/>

            SELECT $this ?feat ?pollutant ?value ?threshold ?unit
            WHERE {
                $this a sosa:Observation ;
                      sosa:observedProperty ?pollutant ;
                      sosa:hasFeatureOfInterest ?feat ;
                      sosa:hasResult/qudt:hasUnit ?unit ;
                      sosa:hasResult/qudt:numericValue ?value ;
                      sosa:hasProcedure xemo:HourlyMean .

                ?t xemo:isThresholdOfPollutant ?pollutant ;
                   xemo:hasThresholdKind xemo:LimitValue ;
                   xemo:hasAveragingPeriod/time:days 1 ;
                   xemo:hasQuantityValue/qudt:numericValue ?threshold .

                FILTER(?value >= ?threshold && ?value < 1.5 * ?threshold)
            }
        """ ;
    ] .
```

Listing A.4: SHACL shape for orange warning detection based on the hourly PM indicator.

```
@prefix sh:    <http://www.w3.org/ns/shacl#> .
@prefix time:  <http://www.w3.org/2006/time#> .
@prefix xemo:  <https://example.org/xemo#> .
@prefix demo:  <https://example.org/xemo/demo#> .
@prefix sosa:  <http://www.w3.org/ns/sosa/> .
@prefix qudt:  <http://qudt.org/schema/qudt/> .

demo:OrangeWarningHourlyShape
    a sh:NodeShape ;
    sh:targetClass sosa:Observation ;
    sh:severity sh:Violation ;
    sh:sparql [
        a sh:SPARQLConstraint ;
        sh:message "Orange warning: Immediate action is required, including temporary suspension of high
dust-generating activities and implementation of intensified mitigation measures." ;
        sh:select """
            PREFIX time: <http://www.w3.org/2006/time#>
            PREFIX xemo: <https://example.org/xemo#>
            PREFIX sosa: <http://www.w3.org/ns/sosa/>
            PREFIX qudt: <http://qudt.org/schema/qudt/>

            SELECT $this ?feat ?pollutant ?value ?threshold ?unit
            WHERE {
                $this a sosa:Observation ;
                      sosa:observedProperty ?pollutant ;
                      sosa:hasFeatureOfInterest ?feat ;
                      sosa:hasResult/qudt:hasUnit ?unit ;
                      sosa:hasResult/qudt:numericValue ?value ;
                      sosa:hasProcedure xemo:HourlyMean .
```

```
            ?t xemo:isThresholdOfPollutant ?pollutant ;
               xemo:hasThresholdKind xemo:LimitValue ;
               xemo:hasAveragingPeriod/time:days 1 ;
               xemo:hasQuantityValue/qudt:numericValue ?threshold .

            FILTER(?value >= 1.5 * ?threshold)
        }
    """ ;
] .
```