On A Semantic Model and Knowledge Graph Based Approach to Enable Transparency, Explainability, and Auditability for High-Pressure Die-Casting

Stefan Bischof[1,*], Erwin Filtz[1], Josiane Xavier Parreira[1], Florian Rötzer[2], Simon Steyskal[1] and Stephan Strommer[2]

[1]*Siemens AG Österreich, Vienna, Austria*

[2]*AIT Austrian Institute of Technology GmbH, Center for Vision, Automation & Control, Vienna, Austria*

Abstract

This paper addresses the critical challenge of fragmented data and knowledge in high-pressure die-casting environments, where the lack of integrated information hampers effective troubleshooting and compliance with emerging transparency requirements. We developed a comprehensive semantic model that integrates distributed data sources and expert knowledge into a unified knowledge graph framework, explicitly connecting manufacturing processes, failures, metrics, and countermeasures through formalized semantic relationships. Our implementation shows how the resulting architecture successfully transforms traditionally siloed industrial data into an interconnected knowledge representation that distinguishes between specified expert knowledge and actual operational data, enabling systematic reasoning about cause-effect relationships throughout the manufacturing process. The approach provides significant value by enhancing manufacturing transparency and decision support while aligning with Industry 5.0 principles and emerging regulatory frameworks for explainable industrial systems, ultimately supporting more sustainable and efficient manufacturing processes.

Keywords

High-Pressure Die-Casting, Industry 5.0, Industrial Knowledge Graph, Digital Twin

1. Introduction

The convergence of Information Technology (IT) and Operational Technology (OT) is transforming industrial automation systems into complex, distributed cyber-physical systems (CPS). This transformation, coupled with the accelerating adoption of machine learning and AI technologies, creates an urgent need for transparency and explainability; principles that align with both emerging regulatory frameworks like the EU AI Act[1] and the fundamental requirements for operator trust in industrial environments.

High-pressure die-casting (HPDC) represents a critical manufacturing domain where these challenges are particularly evident. In HPDC operations, process data is typically fragmented across machine-specific PLCs (Programmable Logic Controller), proprietary storage systems, and industrial communication protocols such as OPC UA (Open Platform Communications Unified Architecture). Meanwhile, essential expert knowledge exists in various forms but is scattered across documentation, embedded in control systems, or retained as tacit knowledge by experienced operators and engineers. This distribution of both data and knowledge creates significant barriers to systematic problem-solving and troubleshooting.

In HPDC, several aggregates are used to execute specific processes to produce die-cast parts. A dosing unit provides molten metal from a holding furnace to the shot sleeve of a die-casting machine. A piston transports the molten metal into the mould (negative form of the part), while a temperature control unit

[1]http://data.europa.eu/eli/reg/2024/1689/oj

continuously controls its temperature. In the tempered mould, the solidification of the metal occurs. After each shot, a robot picks the solidified part, puts it into a water bath for cooling, lets a punching machine remove sprues, and finally hands the part over for manual inspection. For quality assurance, random samples are analysed by X-ray, CT, and coordinate measuring devices. Between die-casting shots, a spraying unit applies a release agent.

HPDC is a defect-inducing process due the extreme conditions. When failures in materials, products, processes, or aggregates occur, identifying their root causes and determining appropriate countermeasures becomes exceedingly difficult. Experts are required to manually collect and interpret data from disparate systems, often drawing on tacit knowledge that may be neither formally documented nor consistently accessible. This approach is time-consuming, resource-intensive, and heavily dependent on individual knowledge that risks being lost through personnel changes.

In this short paper, we present ongoing work from the DG Assist[2] project, which aims to develop a human-centered assistance system for HPDC that embodies Industry 5.0 principles by simultaneously addressing human needs ('experts in the loop') and sustainability goals ('zero-defect manufacturing'). We propose a comprehensive semantic model that integrates: (a) Process knowledge and machine settings along with equipment (aggregates) and production documentation; (b) Multi-modal data from diverse data sources such as PLCs, X-Ray, MES (Manufacturing Execution System) and results of the quality inspection; and (c) Tacit expert knowledge regarding causal relationships between failures, metrics, and countermeasures.

Our approach leverages semantic web technologies and knowledge graphs to transform traditionally opaque, multi-modal data into transparent, interlinked information that enables explainable recommendations. By creating explicit connections between failures, metrics, and countermeasures, we enable both human operators and automated systems to trace the reasoning behind specific recommendations.

The remainder of this paper is organised as follows: Section 2 reviews related semantic approaches in manufacturing contexts. Section 3 describes our proposed semantic model architecture, initial implementation results, and challenges. Section 4 discusses how the proposed framework enables and supports transparency, explainability, and auditability. Section 5 discusses future work and broader implications for industrial transparency.

2. Related Work

Buchgeher et al.'s systematic literature review [1] examines 24 primary studies on knowledge graphs in manufacturing and production. Their work reveals this as an emerging research area, with most knowledge graphs constructed using top-down approaches and modelled as RDF graphs. The study identifies key use cases including knowledge fusion, creation of digital twins, automated process integration, and code generation. It also highlights several open challenges: better handling of numeric data, enabling real-time knowledge graph updates, and establishing best practices for knowledge graph development in manufacturing.

Mörzinger et al. [2] address a fundamental challenge in manufacturing data access—database queries typically require IT specialists due to complex schemas and technical access methods. Their large-scale framework employs semantic web technologies and ontologies to improve data access, enabling domain experts to formulate queries more intuitively across heterogeneous data sources without extensive IT support.

Recent advances in connecting knowledge graphs with time series data offer promising solutions for manufacturing environments. Bakken and Soylu [3] introduce a hybrid query engine that enables unified querying of knowledge graphs and time series data. Their system connects semantic data in SPARQL endpoints with high-volume time series data stored in various backends (including OPC UA, SQL databases, and cloud services). This allows users to execute portable queries across disparate systems without materialising entire datasets as RDF triples.

---

[2]DG Assist: https://projekte.ffg.at/projekt/4641766
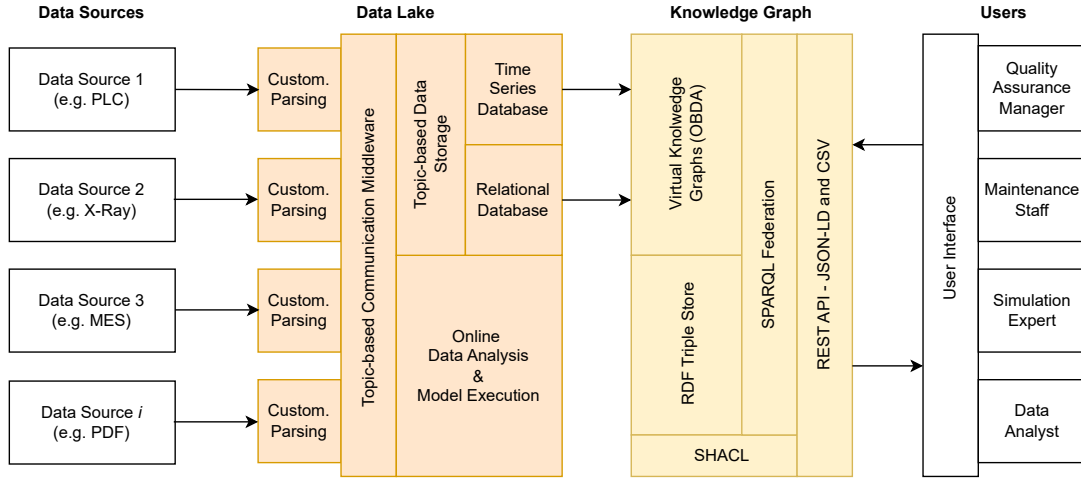
Figure 1: Architecture of the DG Assist data pipeline

Complementing this approach, Meyers et al. [4] present a knowledge graph framework for manufacturing analytics that integrates heterogeneous data sources into a unified semantic representation. Their framework supports ad hoc analysis by making relationships between manufacturing entities explicit and queryable, providing specialised tools for data scientists and knowledge engineers.

The manufacturing process modelling domain has seen significant semantic developments. Kessler and Perzylo [5] present an approach using OWL ontologies to create abstract process models that can be flexibly mapped to specific robot workcells through a Knowledge Base component. Jeleniewski et al. [6] focus on parameter interdependencies, using the DIN EN 61360 standard and OpenMath ontology to formally capture relationships between process parameters at various levels of abstraction.

While these approaches [2, 3, 4, 5, 6] effectively address aspects of data integration and process modelling, our HPDC semantic model advances the state of the art through three key innovations: (1) implementing a comprehensive domain-specific knowledge representation for high-pressure die-casting that formalises causal relationships between process parameters and quality outcomes; (2) integrating tacit expert knowledge with empirical data; and (3) providing role-based explainability mechanisms tailored to different stakeholders. This domain-specific approach creates a more complete solution for manufacturing transparency and auditability that aligns with industrial requirements.

3. Proposed Semantic Model Architecture

This section presents both our proposed semantic model architecture and its practical implementation within an operational HPDC environment. We introduce the conceptual framework designed to address the core challenges of data and knowledge fragmentation in manufacturing settings, then progress to its specific realisation in the DG Assist project with a medium-sized manufacturer of die-cast parts.

Overall Structure of the Semantic Model and Primary Components The layered architecture of our approach (see Fig. 1) comprises the (disparate) data sources, the data lake, knowledge graph foundation including the semantic model, and the interface components. *Data Sources* include sensor data via PLCs, different types of images (X-Ray, CT, potentially photos from thermal cameras, data from the MES, documents, and data which is manually collected. The data of all of these sources are parsed and integrated into the *Data Lake*. Data is represented and stored fitting its purpose and characteristics. For example, sensor data is stored in a timeseries database, which can ingest data efficiently. Other data might be stored in relational databases, or in RDF triple stores. The *Knowledge Graph* of the system links the entities represented in the data lake together to one federated RDF-based knowledge graph, using (materialised) triple stores and federated knowledge graphs with SPARQL 1.1 Federation.
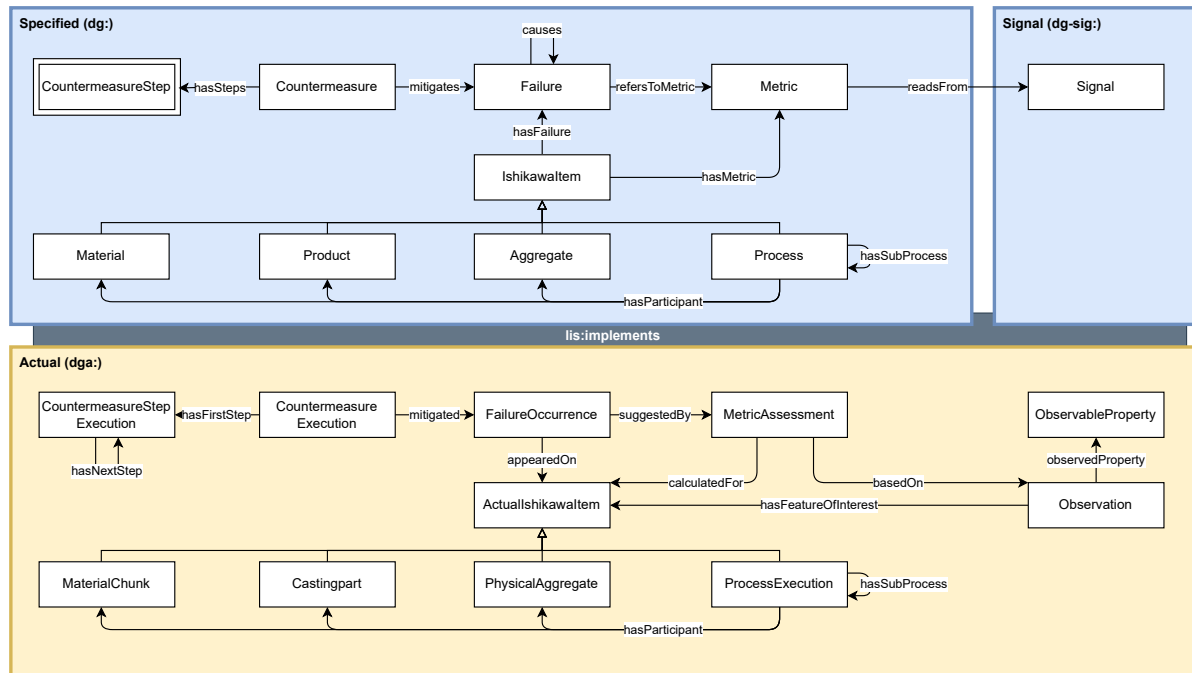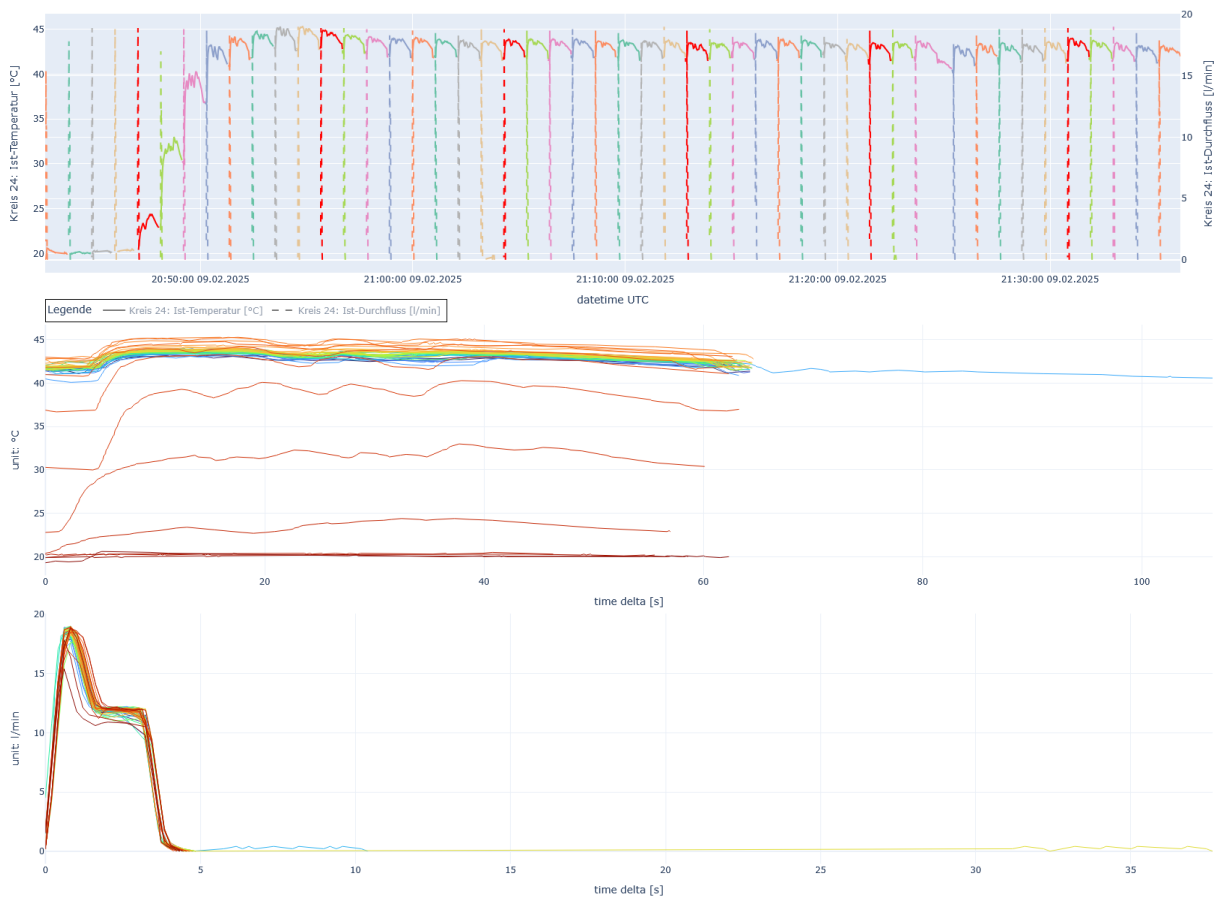
Figure 2: Overview of the HPDC semantic model

Ontological Framework The semantic model must accommodate a large variety of data, including the data captured by the DG Assist data pipeline, as described above. We build the HPDC semantic model by following the Linked Open Terms methodology.[3] During the ontology implementation phase we used and extended several existing ontologies playing different roles in the semantic model: Industrial Data Ontology (IDO) as an upper ontology, Dublin Core (DCT) for term annotations, Semantic Sensor Network Ontology (SOSA/SSN) for timeseries data, Procedural Knowledge Ontology (PKO) for countermeasures[4].

Figure 2 shows the main classes and object properties of our HPDC semantic model in our federated knowledge graph. The two large blocks on top (blue, 'Specified (dg:)') and on the bottom (yellow, 'Actual (dga:)') distinguish expert knowledge, which describes how certain entities are specified, from actual data, which is gathered during the actual production. This can also be seen as the distinction between engineering data/knowledge and operational data/knowledge. For example, in the `Specified` part of the knowledge graph we represent which product the company is contracted to produce with all of the information connected to it, and in the `Actual` part of the knowledge graph we represent actually produced casting parts with their context. Connecting the corresponding entities between these parts enables reasoning about cause-effect relationships, that ranges from concrete casting parts and their failures all the way to process knowledge of available countermeasures.

Mechanisms for Capturing and Formalising Expert Knowledge We examine our approach to transforming tacit expert knowledge into explicit, machine-interpretable representations. This subsection covers our knowledge acquisition method, focusing on formalizing causal relationships between failures, metrics, and countermeasures that create transparent reasoning chains supporting explainable recommendations. (a) *Production facility:* Based on Ishikawa Analyses, we gathered the main concepts of HPDC (processes, equipment, materials) and explored their interconnections in expert interviews. (b) *Root causes:* Based on FMEA (Failure Mode and Effects Analysis) and 5-Why Analyses, we held expert workshops to trace the causalities of product quality issues all the way

---

[3]https://lot.linkeddata.es/
[4]see https://rds.posccaesar.org/ontology/lis14/ont/core/, https://www.dublincore.org/specifications/dublin-core/dcmi-terms/, https://www.w3.org/TR/vocab-ssn/, and https://w3id.org/pko

Figure 3: Example report; Water temperature and flow in a cooling circuit of the HPDC system; The time series data is segmented based on separate process executions for separate cast parts.

back to their known root causes. Also, the causes were linked to at least one Item from the Ishikawa Analyses. Causes can be intricately intertwined and build a complex network that is hard to visually represent. For this reason, in collaboration with the HPDC experts, we developed a compartmentalised representation of the causes network, enabling us to keep the overview over an otherwise very complex topic. (c) *Countermeasures:* Further inspired by FMEA, we encouraged the experts to document and attach any known countermeasures to the nodes of the causes network. (d) *Metrics:* Inspired by SPC (Statistical Process Control), for any failures (causes) that are connected to quantifiable data, the experts defined metrics that verify whether a given failure has taken place. These metrics are further connected to the data and signals provided by the data acquisition layer of our solution. They therefore represent the connection between the quantitative and the semantic representations of the production facility.

This knowledge acquisition method is currently based on human interaction. However, with the process in place, it might be possible to further formalise and automate it in the future. Note, however, that one important prerequisite, which was not given in the project, is that the necessary documents (Ishikawa, 5-Why, FMEA, SPC) can be used and made sense of without an expert's explanation. To this end, it is necessary to properly formalise the processes and formats for these documents and semantically enrich them to make them machine-readable.

Technical Implementation and Deployment   Figure 3 shows an example of a preliminary ad-hoc analysis (split into three parts) that uses `Process` descriptions to segment time series data by their separate `ProcessExecutions`, i.e. individual instances of the process, as they occur over time. Currently, plots like these are delivered, for entire shifts, once per day, where the top most plot shows the temperature and flow time series over time, with each process run colored distinctly. The two plots below show the two time series individually but with the x-axis representing the relative time delta

for each process execution instead of absolute time. The plots are gathered in an HTML report file for easy access. With these reports, an expert or worker in charge can easily spot anomalies in the production process, or even go ahead and define a `Metric` and `Failure` node to automatically trigger a `Countermeasure`.

4. Enabling Transparency, Explainability and Auditability

This section examines how our semantic architecture directly addresses the requirements for transparent, explainable, and auditable industrial assistance systems. Unlike conventional 'black box' approaches, our model creates transparency through its explicit representation of knowledge as interconnected, queryable relationships. We analyse how the knowledge graph structure provides natural transparency by mirroring human understanding of HPDC. We then examine the explainability mechanisms that generate contextually appropriate explanations tailored to different user roles. Finally, we assess auditability features including provenance tracking, reasoning pathway preservation, and confidence attribution. Throughout, we demonstrate how these capabilities enhance human-machine collaboration in manufacturing environments where process understanding and decision justification are operationally critical.

4.1. Transparency Through Knowledge Graph Representation

As highlighted in the previous section, data related to the die-casting process is highly heterogenous, both in terms of the nature of the data (e.g., static vs. dynamic) but also how data is stored (time series vs. relational databases). This setup imposes many challenges in finding relevant data for a particular application of analysis. It is often the case that such data silos are maintained by different people, and an overview of the entire die-casting process is often difficult to obtain.

Our semantic modelling and architecture generates transparency by taking disconnected data sources and creating an interconnected semantic framework. By integrating expert data, sensor readings, and quality inspection results into a unified knowledge representation, data can be easily found, and the relationship between data and the different steps of the die-casting process is clear.

Moreover, the graph-based representation of our approach inherently supports transparency as it aligns with how domain experts naturally conceptualise die-casting processes, creating intrinsic transparency that mirrors human mental models. Unlike traditional database approaches that hide relationships in abstract schemas, our graph representation makes connections between entities explicit and immediately visible. Each node in the graph represents a well-defined manufacturing concept (e.g., process parameters, equipment components, quality metrics), while edges represent clear semantic relationships between these concepts. For instance, when a failure occurs, the knowledge graph reveals not just the affected component but its connections to relevant metrics, thus providing immediate contextual awareness that traditional systems lack.

In summary, our semantic architecture supports transparency at multiple layers, addressing different stakeholder needs: (a) *Data source transparency:* The model explicitly represents the origin of each data element through source attribution metadata, allowing users to trace any information back to its primary source whether from PLC systems, MES, or quality inspection systems. (b) *Semantic transparency:* Domain concepts are explicitly defined with formal semantics rather than implicit encodings, enabling both humans and machines to understand the meaning of data without specialised translation. (c) *Process transparency:* The causal relationships between failures, metrics, and countermeasures, are made explicit through semantic links that can be visualised and queried. (d) *Decision transparency:* When the system suggests countermeasures, the knowledge graph allows tracing backward through the reasoning chain to understand why specific recommendations were made.

4.2. Explainability Mechanisms

As shown in Figure 1, there are many different data sources involved in a HPDC process, including documentation in text format, data from PLCs and a MES as well as (x-ray or thermal) images taken from the actual manufactured part. All these data sources contain data in their own data format that is produced throughout the HPDC and can be further used for a more detailed analysis. This provides insights in the machine settings and can be used to track down failures in the manufactured parts. Unfortunately, the data comes in different variety and veracity and hinders easy integration. Furthermore, the analysis also involves expert knowledge, which is not persisted in a database.

We argue that a semantic model is crucial to enable explainability in a HDPC process. Explainability in this context means user-tailored explanations based on different user-roles, for instance to technical and non-technical users. Our semantic model, as shown in Figure 2, helps to overcome the challenges traditional manufacturing systems typically lack.

Explicit causal relationships   In traditional manufacturing settings, the failure detection heavily relies on implicit knowledge about failures and their potential causes. In our semantic model, we make the relations between `dg:Failure` and a `dg:Metric` visible by explicitly describing which failure is detected by which metric. Additionally, each failure can be mitigated by one or more `dg:CounterMeasures` with explicitly described steps of what needs to be done in order to resolve the failure or to prevent future failures.

Integration of expert knowledge with empirical data   Data collected throughout a manufacturing process can provide insights about the deviations from the operating standards. Purely data-driven approaches are prone to miss expert knowledge acquired over years and experience. However, expert knowledge is often implicit and not persisted in a structured and manageable way. Making knowledge explicit is highly beneficial for industry as knowledge stays within the company in case employees leave, it can be used for training of new employees, is standardized and can be managed. The full potential of failure resolution and applying countermeasures can be only unfolded when expert knowledge and empirical data from the manufacturing process are combined and persisted.

By addressing the challenges outlined above, the manufacturing process benefits from structured and traceable mitigation activities increasing the overall production quality as it is based on the combination of sensor data and expert knowledge stored in a unified model. The granularity of explanations can also be adjusted to different user roles, for instance technical or non-technical users.

4.3. Auditability Features for Industrial Compliance

Instance data can be used to audit the die-casting process by leveraging the semantic model that distinguishes between specified process elements and their actual executions (cf. Figure 2).

Instance data captured during production provides quantifiable metrics for auditing through the `MetricAssessment` concept. Specifically, process performance can be evaluated by analyzing `MetricAssessment` instances related to `ProcessExecution` entities. Signal analysis is enabled by linking `Signal` measurements with production outcomes through `ObservableProperty` instances. Furthermore, quality tracking is achieved by monitoring `FailureOccurrence` instances and their relationship to `MetricAssessment` values.

Listing 1 demonstrates how metric assessments can be represented in a concrete porosity defect scenario.

The model supports comprehensive failure analysis by explicitly linking `FailureOccurrence` instances to the underlying `MetricAssessment` data that suggested them, enabling systematic identification of failures. Root cause analysis is facilitated through the use of `IshikawaItem` concepts, which categorise and analyse failures within the semantic model. Furthermore, `CountermeasureExecution` instances indicate applied countermeasures for specific `FailureOccurrence` events, allowing for detailed tracking and evaluation of countermeasures over time.

Auditability Mechanisms Built on Explainability Our explainability framework enables explanations based on user roles and expertise levels with the manufacturing knowledge graph. This adaptation ensures that explanations are both accessible and actionable for different stakeholders: *Operator-oriented explanations* focus on actionable insights and immediate corrective measures, highlighting machine settings that can be directly adjusted and their expected effects on quality outcomes. *Engineering-oriented explanations* provide deeper technical detail, enabling a wide range of standardised and ad-hoc analyses, including statistical analysis or historical patterns.

Our explainability mechanisms are inherently designed with audit-

```
# Process execution instance
ex:process-123 a dga:ProcessExecution ;
    dga:hasParticipant ex:material-123 .

# Observable property for this process
ex:observable-123 a dga:ObservableProperty ;
    dga:observedProperty <https://w3id.org/dgassist/signal/energy> .

# Metric assessment based on the observable property
ex:metric-123 a dga:MetricAssessment ;
    dga:calculatedFor ex:process-123 ;
    dga:basedOn ex:observable-123 ;
    dga:value "4.82"^^xsd:decimal .

# Failure occurrence instance
ex:failure-456 a dga:FailureOccurrence ;
    dga:suggestedBy ex:metric-456 ;
    dga:appearedOn ex:material-123 .

# Countermeasure execution linked to failure
ex:countermeasure-789 a dga:CountermeasureExecution ;
    dga:mitigated ex:failure-456 ;
    dga:hasFirstStep ex:step-1 .
```

Listing 1: Audit trail example

ability principles in mind, ensuring that explanations can be verified and traced to their underlying evidence. While explainability focuses on making the system's reasoning understandable to users, these same semantic structures simultaneously build the foundation for systematic verification. The semantic model supports this dual purpose through: (a) *Traceable reasoning paths:* Each explanation generated by the system is represented as a formal path through the knowledge graph that can be preserved, examined, and verified against source data. These reasoning paths capture the full provenance of how conclusions were reached. (b) *Evidence linking:* Explanations explicitly connect conclusions to supporting evidence, whether that evidence comes from sensor measurements, quality inspections, or encoded expert knowledge. This linking enables verification of the factual basis behind explanations. (c) *Decision provenance:* When countermeasures are recommended as part of an explanation, the system maintains explicit links to the reasoning that led to those recommendations, creating natural accountability.

For instance, the system processes sensor measurements (such as temperature, pressure, and flow rates) and links it to the corresponding `ObservableProperty` instances. Combined with the definitions of `Metrics`, these properties enable the creation of `MetricAssessments` with specific values. When these values meet certain conditions defined within the `Metric`, the system generates a `FailureOccurrence` (such as porosity in a casting part) and links it to the corresponding expert knowledge `Failure` definition. The system then determines possible root causes based on these connections and recommends appropriate `CounterMeasures` from the knowledge base. These recommendations are presented to the personnel, who may choose to execute them, with the resulting `CountermeasureExecution` instance completing an auditable trail from initial sensor readings through failure detection to resolution actions.

By leveraging the semantic model's distinction between specified and actual elements, our architecture enables comprehensive auditing capabilities. This includes the ability to compare planned processes (`Process`) with their actual executions (`ProcessExecution`), analyzing the effectiveness of specified countermeasures by evaluating their real-world implementations (`Countermeasure` versus `CountermeasureExecution`), tracking material and product entities (`MaterialChunk` and `CastingPart`) throughout their lifecycle, and correlating defined signals (`Signal`) with their instantiated observable properties (`ObservableProperty`). These capabilities collectively support detailed troubleshooting and quality assurance verification that satisfies both operational requirements and regulatory compliance needs.

5. Conclusions & Outlook

This paper has presented our ongoing work on a semantic model and knowledge graph approach for high-pressure die-casting that enhances transparency, explainability, and auditability in manufacturing environments. By integrating fragmented data sources and dispersed expert knowledge into a unified semantic framework, we create explicit connections between manufacturing processes, failures, metrics, and countermeasures. Our approach transforms traditionally siloed industrial data into interlinked knowledge that enables transparent decision-making and systematic troubleshooting. The semantic architecture distinguishes between specified expert knowledge and actual operational data, supporting comprehensive process optimization while aligning with Industry 5.0 principles. Despite these advantages, our approach currently requires significant manual knowledge acquisition efforts from domain experts and faces integration challenges with legacy manufacturing systems that lack standardised data interfaces.

Future work includes finalising the semantic model through continued refinement of the HPDC domain concepts and relationships. We will continue acquiring expert knowledge to populate the knowledge graph with additional failure modes, metrics, and countermeasures relevant to industrial die-casting operations. Implementation of the causal reasoning mechanisms will be a key focus, moving beyond the current preliminary analyses to more comprehensive diagnostic capabilities. We plan to conduct thorough evaluations in production environments to assess the practical benefits of our approach under real-world conditions. The development of human-machine interfaces will is going on to better support different user roles within manufacturing organisations, ensuring that complex semantic insights are presented in accessible ways.

Acknowledgments

Declaration on Generative AI

During the preparation of this work, the authors used Sonnet 4.0 in order to: paraphrase and reword, improve writing style, and grammar and check spelling. After using these tool, the authors reviewed and edited the content as needed and take full responsibility for the publication's content.

References

[1] G. Buchgeher, D. Gabauer, J. Martinez-Gil, L. Ehrlinger, Knowledge graphs in manufacturing and production: A systematic literature review, IEEE Access 9 (2021) 55537–55554. doi:10.1109/ACCESS.2021.3070395.

[2] B. Mörzinger, T. Weiler, T. Trautner, I. Ayatollahi, B. Angerer, B. Kittl, A large-scale framework for storage, access and analysis of time series data in the manufacturing domain, Procedia CIRP 67 (2018) 595–600. doi:10.1016/j.procir.2017.12.267.

[3] M. Bakken, A. Soylu, Chrontext: Portable sparql queries over contextualised time series data in industrial settings, Expert Systems with Applications 226 (2023) 120149. doi:10.1016/j.eswa.2023.120149.

[4] B. Meyers, H. Vangheluwe, P. Lietaert, G. Vanderhulst, J. van Noten, M. Schaffers, D. Maes, K. Gadeyne, Towards a knowledge graph framework for ad hoc analysis in manufacturing, Journal of Intelligent Manufacturing (2024) 1–22. doi:10.1007/s10845-023-02319-6.

[5] I. Kessler, A. Perzylo, Flexible Modeling and Execution of Semantic Manufacturing Processes for Robot Systems, in: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA), 2024, pp. 1–8. doi:10.1109/ETFA61755.2024.10710791.

[6] T. Jeleniewski, J. Reif, F. Gehlhoff, A. Fay, Parameter Interdependencies in Knowledge Graphs for Manufacturing Processes, in: 2024 IEEE 29th International Conference on Emerging Technologies and Factory Automation (ETFA), 2024, pp. 1–8. doi:10.1109/ETFA61755.2024.10711066.