Monitoring and Analytics for the whole Lifecycle, on Models, Hardware, and Software

D5.1 Specification for Semantic Integration and Analysis

Deliverable Due Date	2026-01-31
Deliverable Type	Document
Deliverable Number	D5.1
Authors	Wasif Afzal (MDU, Sweden), Henrik Morast (Alstom, Sweden), Batur Alp Akdoğan (LTG, Turkey), Lajos Schretner (Eli Alps, Hungary), Tóth László (Uni. Szeged), Mattias Mohlin (HCL, Sweden), Mathias Axling (CNET)
Dissemination Level	Public

Contributors

Name	Short Affiliation
Wasif Afzal	MDU, Sweden
Henrik Morast	Alstom, Sweden
Batur Alp Akdoğan	LTG,Turkey
Lajos Schretner	Eli Alps, Hungary
Tóth László	Uni. Szeged, Hungary
Mattias Mohlin	HCL, Sweden
Ercan Ilık	Evosoft,Turkey
Hüseyin Aykanat	BDC,Turkey
Mathias Axling	CNET, Sweden

Reviewers

Name	Short Affiliation	Date
Wasif Afzal	MDU, Sweden	2026-01-15
László Tóth	Uni. of Szeged, Hungary	2026-01-12

Document Revision Log

Version	Revision	Date	Description	Author
01	01	2025-12-01	First Draft	Wasif Afzal (MDU)
02	02	2025-12-05	Revision	Hnerik Morast (Alstom)
03	03	2025-12-05	Revision	Batur Alp Akdoğan (LTG)
04	04	2025-12-16	Revision	Tóth László (Uni. Szeged), Lajos Schretner (Eli-Alps), Mattias Mohlin (HCL)
05	05	2026-01-12	Revision	Wasif Afzal (MDU)
06	06	2026-01-13	Revision	Wasif Afzal (MDU)
07	07	2026-01-14	Revision	Batur Alp Akdoğan (LTG)
08	08	2026-01-16	Revision	Mathias Axling (CNET)
09	09	2026-01-19	Final	Wasif Afzal (MDU)

Table of Contents

List of Figures

List of Tables

1. Executive Summary of Deliverable

This delivery provides an in-depth analysis of WP5 (semantic analysis). Further, it is the basis for the project's implementation of that work package. The objective of this executive section is to inform of the project's results, highlight significant findings, and recommend actionable steps moving forward. Overall, the analysis outlines the two main categories of analysis (root cause analysis and predictive analysis), which shows the versatility of MONA LISA. Further, the summary defines the context in which WP5 interfaces with upstream work packages (WP3, WP4) and visual analytics (WP6) as downstream output.

This deliverable D5.1 lays the foundation of ML and AI-based analysis of requirements and data, coming from diverse sources, and made available by MONA LISA's use case providers. The deliverable lists down the inputs (format, source etc.), the details of the intended analysis and the desired outputs from each use case provider. The deliverable then further gathers the inputs from MONA LISA solution providers, with the aim to summarize the solution offerings, and to pave the way for upcoming solution development in WP5. The deliverable has identified two broad categories of analysis from the input of use case providers: (1) automated root cause analysis of defects and (2) predictive analytics. The solution providers aim to target these analysis categories through their expertise in ML and LLM-based solutions, with a particular emphasis on data extraction, processing and validation. This deliverable ends with a brief recap of state of the art on MONA LISA-relevant analysis.

2. Introduction

The main objective of work package 5 is to provide the semantic layer for MONA LISA's analytics and business logic. This WP receives various data (e.g., code traces and logs) from the infrastructure (WP3) and models from the model reengineering layer (WP4) and performs advanced analyses using methods from the subdomains of Artificial Intelligence (AI), e.g, Machine Learning (ML), Retrieval Augmented Large Language Models (RA LLM) and model-based approaches from WP4. The analysis then presents results for the visual front end (WP6). Figure 1 shows the interactions of various WPs with WP5.

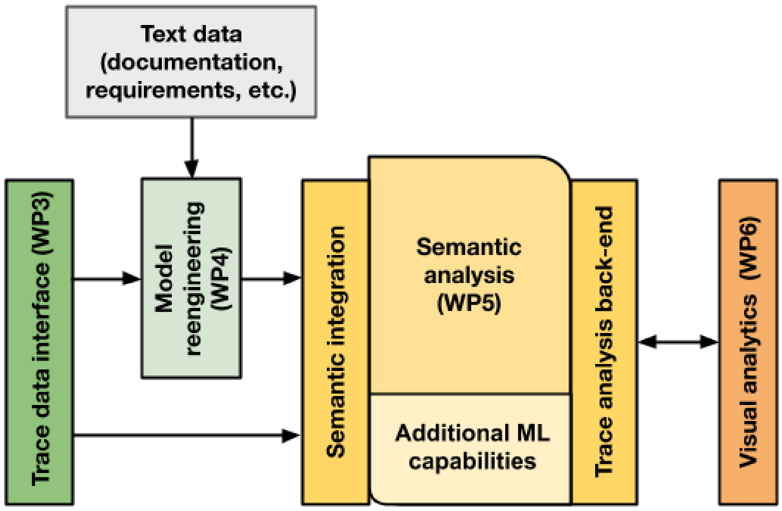


Figure 1. ***Interaction between the analysis work packages (WP4, WP5) and their role within the overall pipeline.***

WP5 has following two objectives:

- *Objective 1:* Semantic data integration using Knowledge Graphs and Language Models. Developers map data sources and analysis tools based on their knowledge of data semantics with assistance from ontology-based data access tools using domain-specific knowledge graphs (KG). Additional automation will be developed through Large Language Models (LLMs).
- *Objective 2:* Data analysis with ML and Retrieval-augmented LLMs. The aim is to integrate the semantically unified data model with both ML and LLM backends and tools. These tools and libraries will provide the analysis capabilities of the MONA LISA solution, giving users new insights into data traces.

The purpose of D5.1 is to define the foundations to enable exchange of data between the use case providers and the solution/knowledge providers towards ML and LLM based data analysis.

3. Inputs, Types of Analyses and Outputs

In this section, we list, for each use case, the inputs, the details of the analyses, and the outputs. The motivation is that this will serve as the basis for defining the semantics of data exchange among WP3/WP4 utilities, the specific analysis toolkit, and the output for WP6.

3.1 Use case domain 1 – Infrastructure, Mobility & Logistics

3.1.1 Alstom

Root cause analysis.

3.1.1.1 Inputs

- Tickets (Business ticketing tool Dr MITRAC, SharePoint predating Dr MITRAC)
    - A submitted product-issue ticket, including attached log files
    - Related tickets, including resolved ones
- Logs (plain text, not structured)
    - safeprime.log - safety-related log
    - hmi_runtime.log - HMI application level log
    - cm_processmonitor.log - start of application log
- Product documentation (Configuration Management tool, Dimensions, SharePoint)
- Source code (GitLab)
- FWIs - Defects, Change Requests and Feature Requests (Engineering Workflow Management system)

3.1.1.2 Type of Analyses

AI-supported analysis using Large Language Models (LLMs) for natural language understanding and Retrieval-Augmented Generation (RAG) to combine ticket/log retrieval with generative reasoning.

3.1.1.3 Outputs

- Log visualisation
- Root cause analysis
    - Identification of causes, through logs, tickets, and documentation
- Solution or mitigation proposals

3.1.2 CNET

Improve estimation of remaining useful life (RUL) of equipment and optimisation of needed maintenance, and improve visualisation techniques, and digital twin development.

3.1.2.1 Inputs

- Structured output (JSON)
    - Measurements from sensors
    - Device state data
    - Logs

3.1.2.2 Type of Analyses

The use case will apply AI-supported visual analytics, unified trace data storage, and model-based behaviour analytics to enable continuous monitoring and assessment throughout the lifecycle.

3.1.2.3 Outputs

- Log visualisation
- System status analysis
- RUL analysis

3.1.2.4 Architectural Diagram

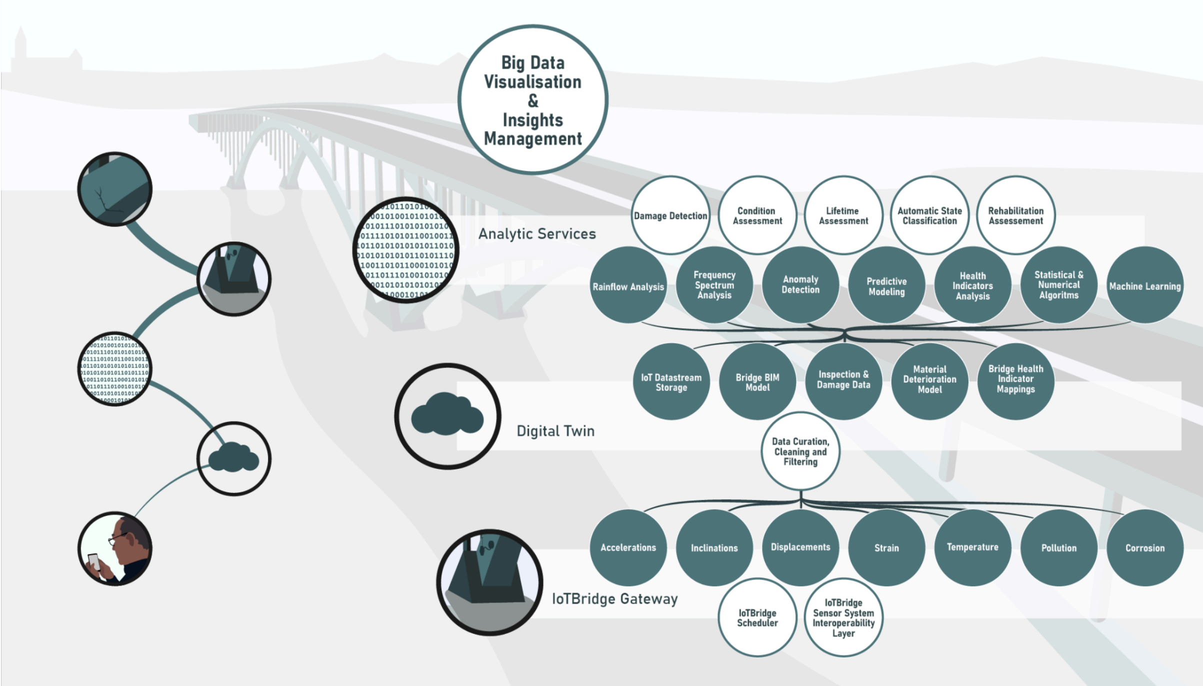


Figure 2. *CNET's use case.*

3.2 Use case domain 2 – High-tech Equipment and Industrial Operations Technology

3.2.1 Eli Alps

3.2.1.1 Inputs

Logs exported into csv format. The logs are well-formed and also human-readable. They contain the event identifier, the event source, and the description.

3.2.1.2 Type of Analyses

- AI-supported analysis
- Root cause analysis
- Error detection and localisation

3.2.1.3 Outputs

Visual analysis to support error localisation and root-cause explanation.

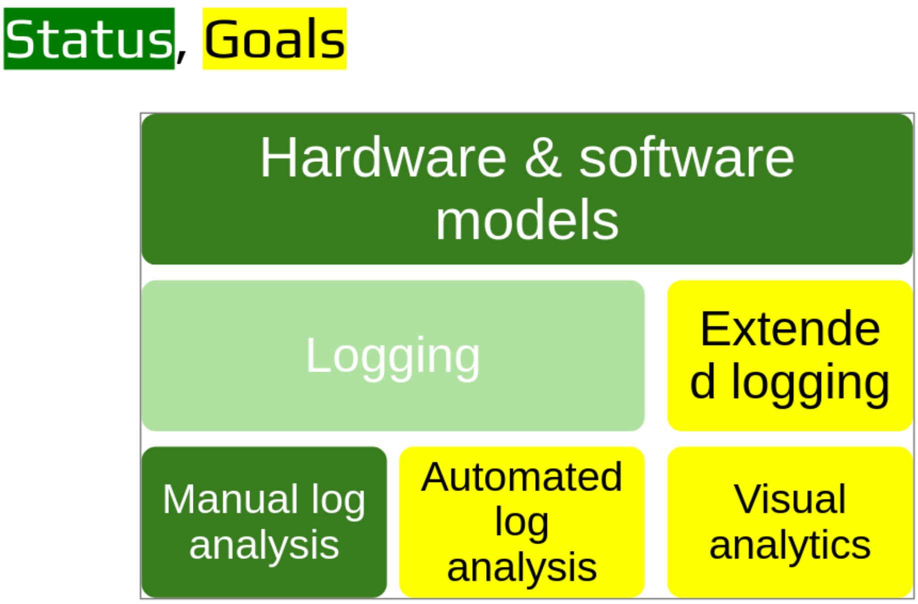


Figure 3. ***Eli Alps' use case for automated log analysis.***

3.2.2 Bilecik Demircelik

3.2.2.1 Inputs

This section defines the data acquisition strategy, analytical framework, and output mechanisms for the Induction Furnace use case at Bilecik Demir Çelik (BDÇ). By utilizing the MONALISA platform, the objective is to shift the facility's maintenance paradigm from a reactive model—where actions are taken solely after failure—to a predictive, data-driven strategy.

The system specifically targets three high-impact operational challenges identified in the domain analysis:

- **Refractory Lifecycle Prediction:** To mitigate the safety risks of molten-metal leakage and optimize the furnace lining replacement schedule, currently managed through imprecise manual measurements.
- **Earth Leakage:** To detect and predict earth leakage failures by correlating ground leakage signals with cooling system anomalies, effectively mitigating risks associated with insulation degradation and water leaks.
- **Code 39 Faults:** To analyze and prevent generic electrical faults that cause undefined system stops, isolating root causes.

The architecture integrates high-frequency sensor telemetry with discrete operator logs to transition from reactive maintenance to predictive analytics.

The system aggregates heterogeneous data streams ranging from high-frequency automated PLC signals to manual operator entries regarding failure modes. Furthermore, the system incorporates raw material inputs regarding scrap composition. This includes data on the type of scrap charged to the furnace, but the input is currently low-granular and lacks a detailed classification structure.

3.2.2.1.1 Sensor Data (Time-Series)

The system leverages a comprehensive array of continuous sensor telemetry acquired directly from the furnace PLCs and persisted in an MS SQL Server environment. Currently, this dataset comprises approximately **116 distinct signal tags** available through existing **SCADA screens**. However, operators do not actively monitor the entire spectrum of data simultaneously; their focus is typically limited to a subset. While the current SCADA visualization primarily serves for real-time observation of these primary metrics, the MONALISA platform uses the full high-dimensional dataset for historical trend analysis and predictive modeling. The data is predominantly encoded as IEEE 754 floating-point numbers for analog measurements and signed 32-bit integers for electrical metrics.

The sensor data landscape covers the following critical categories, each selected for its relevance to the target failure modes:

- **Water Flow and Temperature Data:** The system monitors the entire cooling circuit to generate thermodynamic profiles critical to analyzing **all three target failure modes**. The telemetry includes **Return Water Flow Rates** collected from multiple distinct cooling lines to ensure comprehensive coverage of the furnace geometry. Additionally, the system tracks **Melting Unit Cooling Water Inlet and Outlet Temperatures** to calculate thermal gradients. These hydraulic and thermal metrics are essential for identifying refractory thinning (heat-transfer efficiency), predicting impending water leaks (Earth Leakage), and providing the necessary thermodynamic context to isolate root causes of generic system trips (Code 39).
- **Electrical Parameters:** The system continuously monitors critical electrical parameters, specifically **Phase Currents**, **Phase Voltages**, and power consumption metrics (**Active, Reactive, Apparent Power**). These inputs are indispensable for analyzing **all three target failure modes**, as they reveal electrical imbalances, arcing, and surges indicative of Refractory failure, Earth Leakage, and Code 39 faults.
- **Operational Status Signals:** This category comprises binary flags and discrete-state indicators that provide context for continuous streams. Key signals include the **Ground Leakage Detector**, which directly alerts the system to insulation breaches, and various fault flags associated with operational stops. These signals serve as ground-truth labels for training predictive models to detect earth leakage and system trips.

3.1.1.2.2 Operational Failure Logs (Event Data)

Programmable Logic Controllers (PLCs) accurately identify system stop occurrences and timing, but lack the semantic capability to determine the underlying cause. Consequently, manual operator classification is required to provide context for each downtime event. Operators enter these failure codes via the interface screens shown in **Figure 4**. This discrete-event data is persisted in an MS SQL Server environment that is physically separate from the server hosting the high-frequency sensor telemetry. The system maps these manually entered codes to the three target operational challenges based on historical correlation analysis:

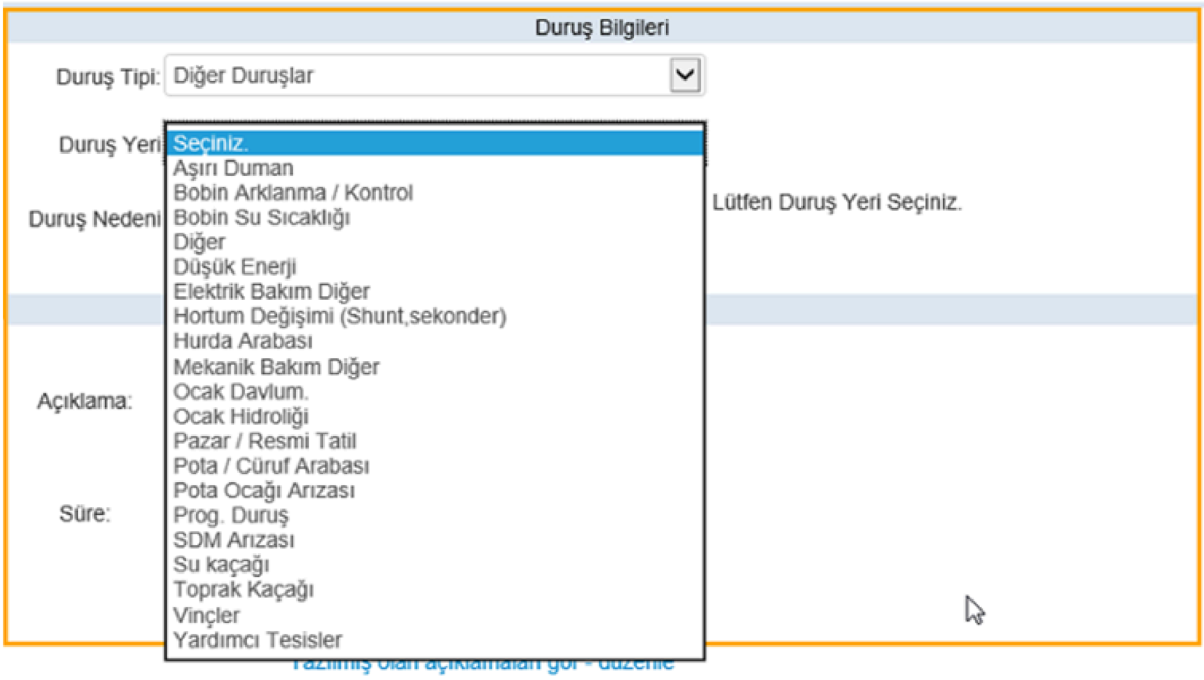


Figure 4. ***Operator downtime login screen.***

Refractory Lifecycle: The system explicitly filters for failure modes that directly and adversely affect the refractory lining's lifespan. These events are categorized as critical stressors that accelerate degradation. Key indicators include structural failures such as **RF-04 (Furnace Burn-through)** and **RF-05 (Defective Lining/Sinter Cracking)**, as well as high-impact electrical events like **RF-01 (Coil Puncture due to Arcing)** and **RF-02 (Internal Coil Arcing)**. Furthermore, the analysis incorporates operational deviations such as **RF-03 (Furnace Overflow)** and **RF-10 (Humidity/Leaks)**, along with water-related failures like **ELG-02 (Coil Water Leakage)** and **ELG-03 (Hose Burst)**, all of which directly compromise the lining's integrity.

Earth Leakage : This category focuses on detecting insulation failures and cooling-system breaches that compromise furnace safety. The log analysis identifies specific failure modes that directly result in ground faults. These include water-system failures such as **ELG-02 (Coil Water Leakage), ELG-03 (Hose Leakage/Burst), and ELG-04 (Faraday Ring & Elbow Leak)**, as well as electrical insulation failures such as **ELG-05 (Coil Burn/Insulation Weakness)**. Furthermore, the dataset includes refractory failures that physically damage the coil isolation, specifically **RF-01 (Coil Puncture due to Arcing)** and **RF-04 (Furnace Burn-through)**, which are critical contributors to Earth Leakage incidents.

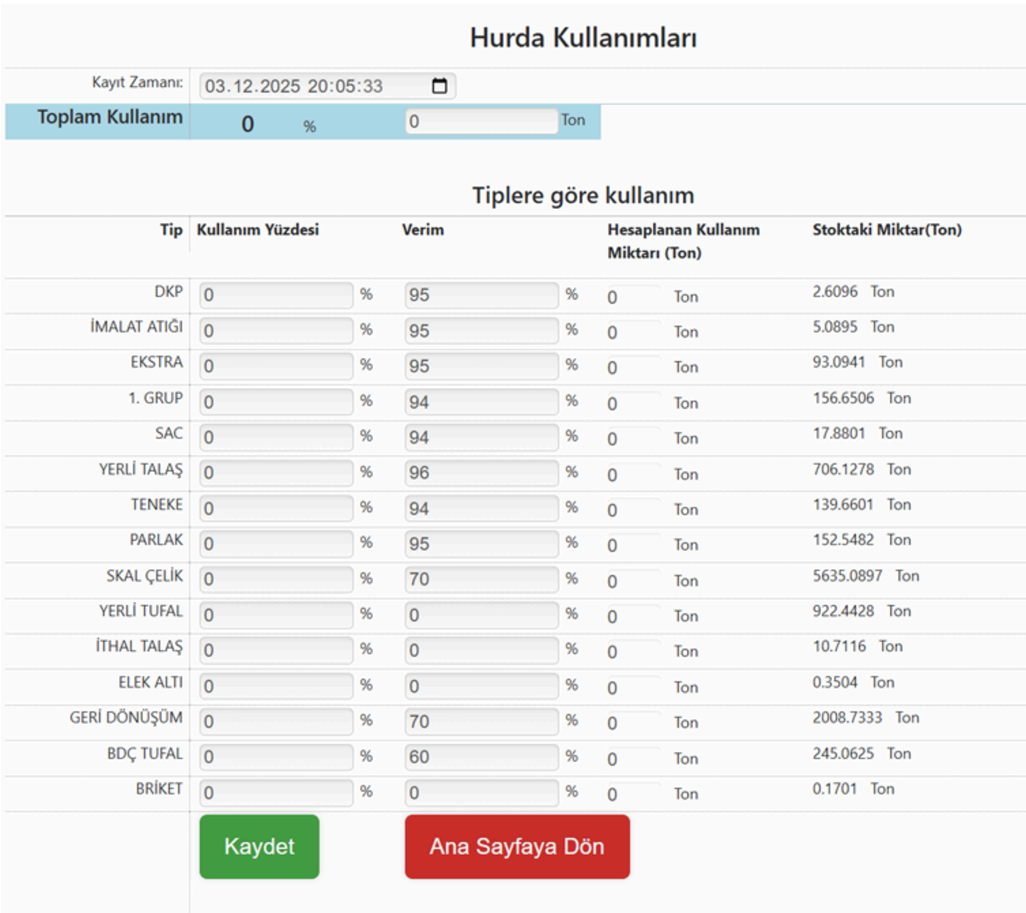
Code 39: While often appearing as a generic electrical fault, "Code 39" is frequently a symptomatic manifestation of critical underlying failures. To disambiguate this signal and isolate the root cause, the analysis explicitly correlates it with specific high-severity events. These include electrical arcing and puncture events such as **RF-01 (Coil Puncture due to Arcing)** and **RF-02 (Internal Coil Arcing)**, as well as operational anomalies like **RF-03 (Furnace Overflow)** and **RF-08 (Loss of Sinter Heating)**. Furthermore, a significant overlap is observed with leakage indicators, including **ELG-01 (Ground Leakage Trip)** and **ELG-02 (Coil Water Leakage)**, confirming that Code 39 often acts as a precursor or secondary indicator for complex system failures.

3.2.2.1.3 Refractory Measurements

Refractory lining thickness is the primary indicator of furnace health. In the current operational workflow, the lining is initially installed using a fixed-geometry mold (template) within the coil, thereby establishing an initial thickness. Actual wear measurements are not performed continuously; instead, manual inspections with a measuring rod are performed only after a specific sequence of casting cycles (heats). This manual method is sporadic and low-precision, yielding only rough depth estimates at a limited set of points. With the project, more regular controls will be initiated, ensuring systematic data acquisition to accurately track degradation trends.

3.2.2.1.4 Raw Material

Scrap Input: Data regarding the raw scrap charged into the furnace. Currently, inputs are categorized into broad classes with low granularity (poor differentiation between scrap types), which introduces variability in the melting process.



Hurda Kullanımları

Kayıt Zamanı: 03.12.2025 20:05:33

Toplam Kullanım 0 % 0 Ton

Tiplere göre kullanım

Tip	Kullanım Yüzdesi	Verim	Hesaplanan Kullanım Miktarı (Ton)	Stoktaki Miktar(Ton)
DKP	0 %	95 %	0 Ton	2.6096 Ton
İMALAT ATIĞI	0 %	95 %	0 Ton	5.0895 Ton
EKSTRA	0 %	95 %	0 Ton	93.0941 Ton
1. GRUP	0 %	94 %	0 Ton	156.6506 Ton
SAC	0 %	94 %	0 Ton	17.8801 Ton
YERLİ TALAŞ	0 %	96 %	0 Ton	706.1278 Ton
TENEKE	0 %	94 %	0 Ton	139.6601 Ton
PARLAK	0 %	95 %	0 Ton	152.5482 Ton
SKAL ÇELİK	0 %	70 %	0 Ton	5635.0897 Ton
YERLİ TUFAL	0 %	0 %	0 Ton	922.4428 Ton
İTHAL TALAŞ	0 %	0 %	0 Ton	10.7116 Ton
ELEK ALTI	0 %	0 %	0 Ton	0.3504 Ton
GERİ DÖNÜŞÜM	0 %	70 %	0 Ton	2008.7333 Ton
BDÇ TUFAL	0 %	60 %	0 Ton	245.0625 Ton
BRİKET	0 %	0 %	0 Ton	0.1701 Ton

Kaydet

Ana Sayfaya Dön

Figure 5. ***Scrap input screen.***

Scrap input entries are provided in Figure 5.

3.2.2.2 Type of Analyses

The analytical framework is structured to serve both immediate operational needs and long-term predictive maintenance goals, categorized as follows:

3.2.2.2.1 Real-Time Descriptive Analysis (SCADA Monitoring)

The SCADA screen, illustrated in **Figure 6**, serves as the primary interface for real-time monitoring. Although the induction furnace infrastructure generates a comprehensive stream of telemetry, not every data point flows through to the central monitoring screens. To enable operators to maintain immediate situational awareness, only a subset of critical operational parameters is tracked in real time.

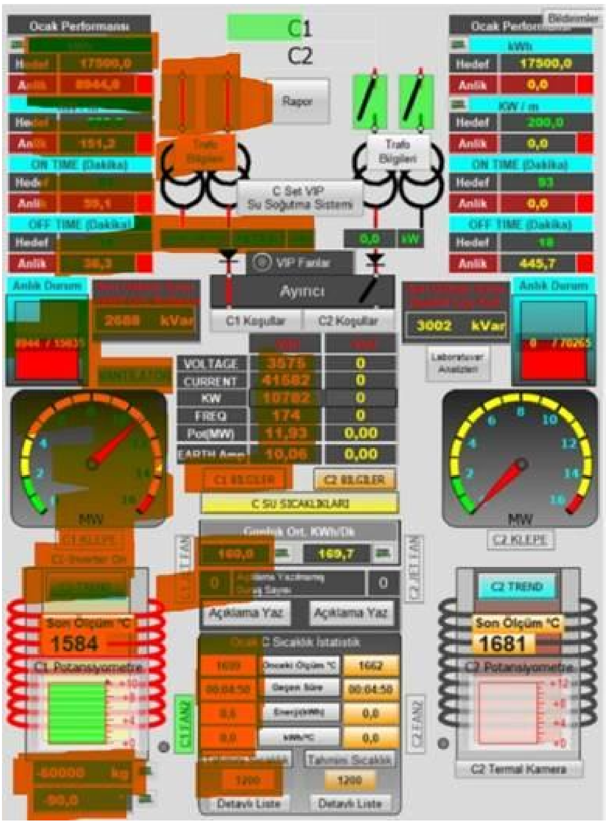


Figure 6. ***SCADA monitoring.***

To facilitate deeper observation, the interface visualizes **trend analysis** using **line graphs** plotted against timestamps. As shown in **Figure 7** this allows operators to observe signal behavior over time rather than relying solely on instantaneous values.

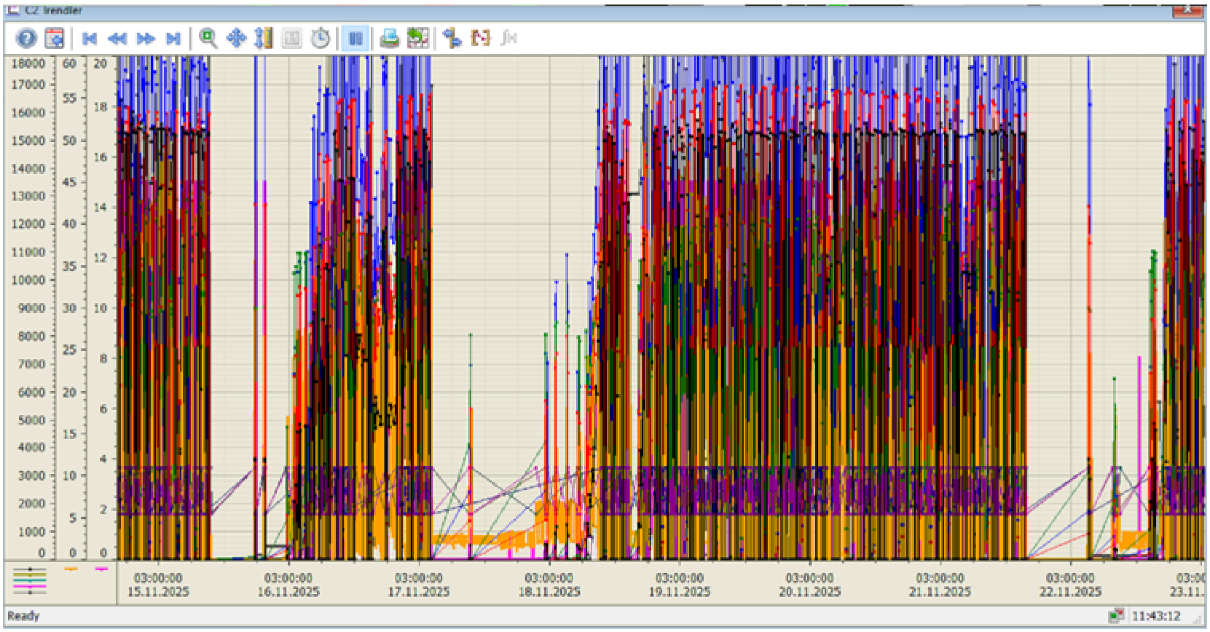
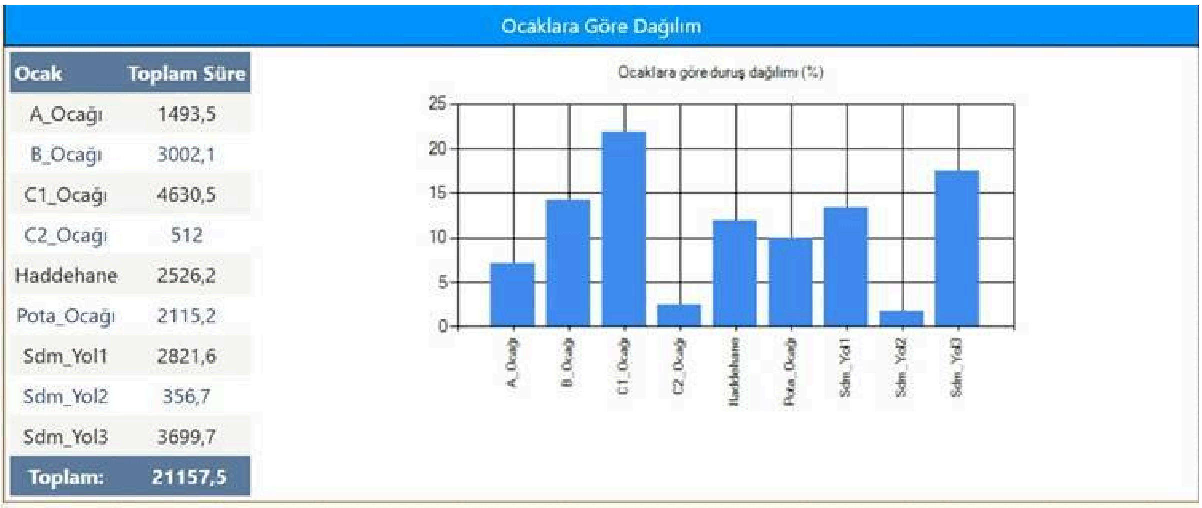


Figure 7. ***Trend analysis.***

3.2.2.2.2 Historical Data Access and Reporting

The separate **.NET-based web server** functions as a repository for simple operational data, specifically the operator-classified failure logs and status inputs. As illustrated in **Figure 8**, the interface generates **summary graphics**, such as the **total downtime per furnace**, enabling management to visually assess key performance metrics. Furthermore, the interface shown in **Figure 9** allows users to filter data by date ranges and export the filtered logs to **Excel** for manual auditing. While the platform supports these basic descriptive visualizations, currently, **no ML or Deep Learning DL algorithms** are deployed for automated anomaly detection or predictive analysis.



Ocak	Toplam Süre
A_Ocağı	1493,5
B_Ocağı	3002,1
C1_Ocağı	4630,5
C2_Ocağı	512
Haddehane	2526,2
Pota_Ocağı	2115,2
Sdm_Yol1	2821,6
Sdm_Yol2	356,7
Sdm_Yol3	3699,7
Toplam:	**21157,5**

Figure 8. ***Website summary graphics.***

BDC
BILECIK DEMİR ÇELİK

Duruş Raporu

Rapor Tarihi: 9.12.2025 14:51:12

Baslangıç:7 Aralık 1.Vardiya Bitiş: 9 Aralık 2.Vardiya

Vardiya Adi	Duruş Saati	Duruş Yeri	Toplam Süre	Birim	Arıza Yeri	Arıza Nedeni	Aciklama	Birim Aciklama
7 Aralık 1.Vardiya	00:04	C2_Ocağı	3,9	Üretim Süreçleri	Cüruf Alma			
	00:04	Pota_Ocağı	4,1	Operasyonel Süreçler	Çalışılan Pota Değişimi			
	00:09	B_Ocağı	4,3	Üretim	Cüruf			

Figure 9. ***Export report.***

3.2.2.2.3 Chemical Composition and Wear Analysis

The system also integrates results from laboratory analyses of the molten metal's chemical composition, which are made accessible through the same **.NET web interface**. This chemical data is analyzed alongside Scrap Input records to identify correlations with **Refractory Thickness**. By linking raw material quality and resulting chemical properties to wear rates, the system aims to identify specific inputs that accelerate lining degradation.

3.2.2.3 Outputs

Upon project completion, our goal is to establish a holistic system that integrates Anomaly Detection and Root Cause Analysis into a fully traceable environment. The key outputs of this system include:

- **Predictive Alerts** derived from Anomaly Detection algorithms that scan high-frequency sensor data to identify irregularities and flag potential issues before system downtime occurs.
- **Automated Root Cause Analysis (RCA) reports** that correlate operational failure logs with sensor telemetry to pinpoint the exact sequence of events and the root cause of the failure.
- **Unified Trace Traceability** on **Trace Compass**, allowing engineers to visualize and track analytical results, signal trends, and event logs on a single, synchronized timeline. This harmonized data will also serve as the foundation for future **LLM-Based Trace Analytics**, where LLM's utilize the middleware-provided data to interpret complex trace files.

4. Tool, Technology & Knowledge Providers' Role

In this section, we capture the capabilities of tool, technology, and knowledge providers to showcase their existing capabilities (infrastructure, data exchange setup, specific data schema, API reuse) or the development of new technologies or capabilities important for achieving WP5 objectives, in terms of the use-case data extracted in Section 3 above. The motivation is to match the tool, technology & knowledge providers' competencies with the data exchange setup and analysis required from the use case providers.

4.1 Evosoft

Evosoft will present information obtained from the middleware to the end user via APIs and transfer trace data to Eclipse Trace Compass in formats agreed upon in WP3, such as Google Trace Event or the Common Trace Format. Data extraction will be managed through APIs tailored to project requirements, such as RESTful APIs, while the middleware infrastructure is planned to use FIWARE or similar technologies. Furthermore, within the scope of WP6, Evosoft is responsible for integrating any additional visualizations required for visual analytics by combining them with data retrieved from the middleware layer, and for developing the platform integrations and tests for the visualization tools.

4.2 Lider

Lider Teknoloji Geliştirme (LTG) will develop a dedicated service to extract SCADA and downtime analysis data currently flowing into MS SQL servers and other databases, and to transfer it to the middleware layer, such as FIWARE. LTG will perform the necessary data harmonization to convert these heterogeneous data streams into a suitable format for the middleware, which will subsequently serve as the primary input for Alpata's ML/DL-based predictive analysis. Additionally, LTG is responsible for ensuring that this data is extracted and processed in real-time to enable immediate visualization and monitoring of the system's status.

4.3 Mälardalen University

4.3.1 DRACONIS (Design Rule Analysis and Checking Of Norms in IEC & Simulink)

DRACONIS is a static analysis framework that automates the review of block-based, safety-critical software, particularly in domains such as railway systems. It automates design rule checks for graphical development tools such as Simulink and Function Block Diagrams (FBD).

The overall toolchain is shown in Figure 10. The current version allows users to input models, which are then parsed into an intermediate representation. Given a model, the analyzer component extracts metrics and performs dataflow analysis to establish a baseline for the analysis results. If any changes are made to the analyzed files, a delta analysis will be triggered, prompting DRACONIS to indicate which checks need to be redone.

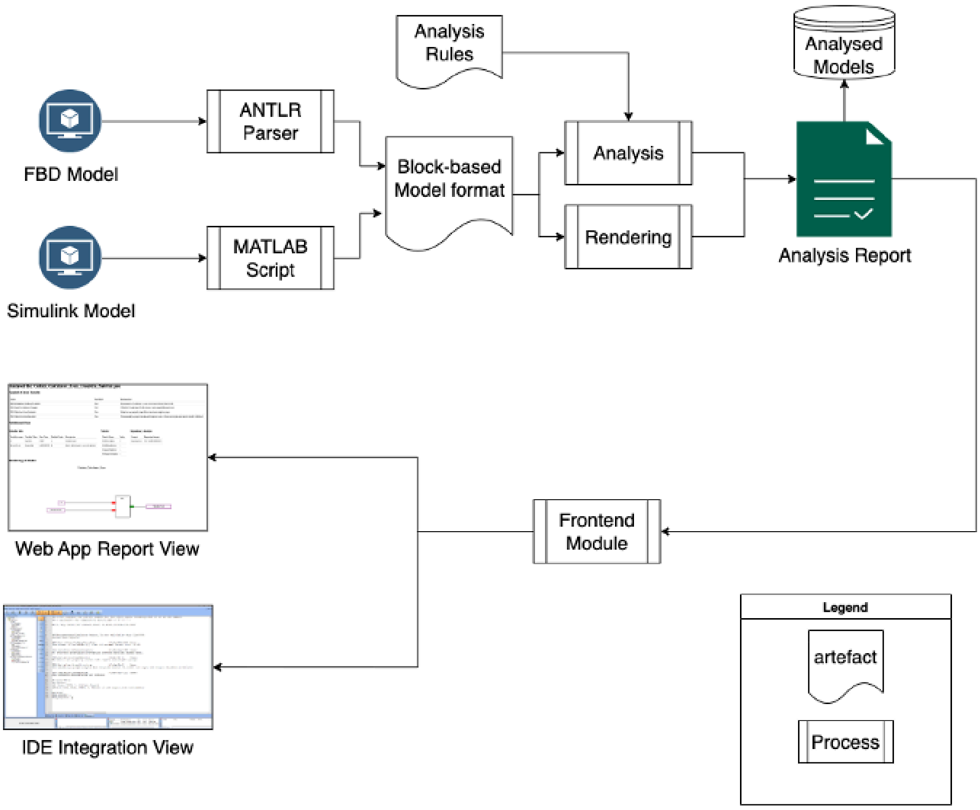


Figure 10. ***DRACONIS toolchain.***

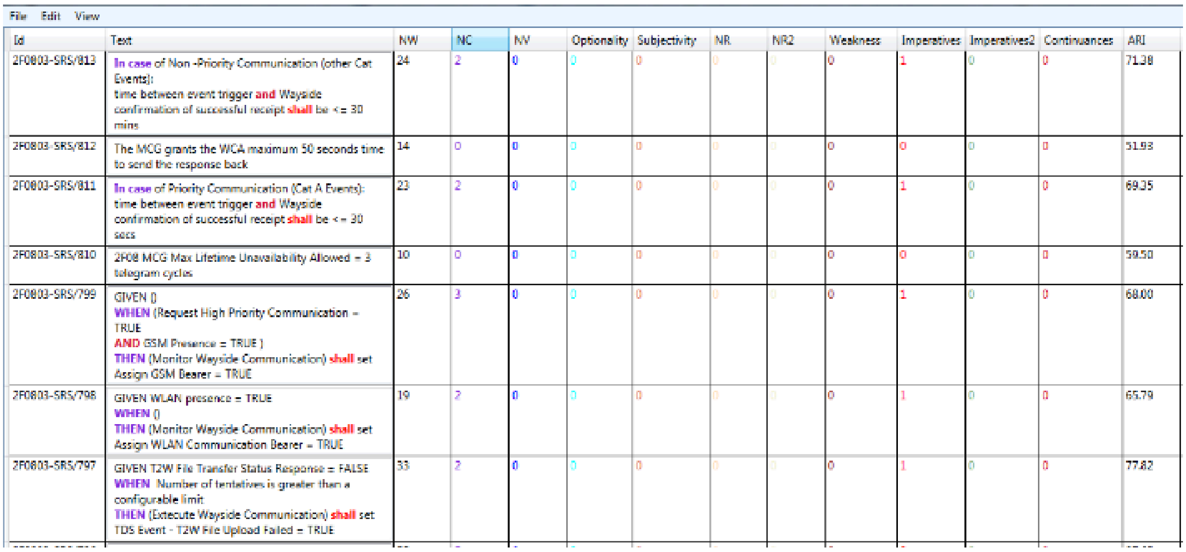
Analysis reports can be delivered in one of two formats: either the tool presents the report in a structured text format, or via the accompanying web application, which also includes a visual model rendering and report review capabilities.

4.3.1.1 Extensions

In the project, MDU will extend DRACONIS with an API frontend and expand its analysis capabilities to support root-cause analysis of failure logs.

4.3.2 NALABS

NALABS processes a collection of documents that include requirements and generates a list of metrics and quality scores for each requirement. This tool is available as a desktop application with a graphical user interface. It requires the requirements to be formatted in an Excel file, with each requirement listed in a separate row. To operate the tool, users must provide information regarding the requirement ID and specification columns. An example of the result view is shown in Figure 11. Metrics are shown in individual columns, and identified issues are highlighted and color-coded by context.



Id	Text	NW	NC	NV	Optionality	Subjectivity	NR	NR2	Weakness	Imperatives	Imperatives2	Continuances	ARI
2F0803-SRS/813	In case of Non -Priority Communication (other Cat Events): time between event trigger and Wayside confirmation of successful receipt shall be <= 30 mins	24	2	0	0	0	0	0	0	1	0	0	71.38
2F0803-SRS/812	The MCG grants the WCA maximum 50 seconds time to send the response back	14	0	0	0	0	0	0	0	0	0	0	51.93
2F0803-SRS/811	In case of Priority Communication (Cat A Events): time between event trigger and Wayside confirmation of successful receipt shall be <= 30 secs	23	2	0	0	0	0	0	0	1	0	0	69.35
2F0803-SRS/810	2F08 MCG Max Lifetime Unavailability Allowed = 3 telegram cycles	10	0	0	0	0	0	0	0	0	0	0	59.50
2F0803-SRS/799	GIVEN () WHEN (Request High Priority Communication = TRUE AND GSM Presence = TRUE) THEN (Monitor Wayside Communication) shall set Assign GSM Bearer = TRUE	26	3	0	0	0	0	0	0	1	0	0	68.00
2F0803-SRS/798	GIVEN WLAN presence = TRUE WHEN () THEN (Monitor Wayside Communication) shall set Assign WLAN Communication Bearer = TRUE	19	2	0	0	0	0	0	0	1	0	0	65.79
2F0803-SRS/797	GIVEN T2W File Transfer Status Response = FALSE WHEN Number of tentatives is greater than a configurable limit THEN (Execute Wayside Communication) shall set TDS Event - T2W File Upload Failed = TRUE	33	2	0	0	0	0	0	0	1	0	0	77.82

Figure 11. ***The GUI interface of NALABS.***

NALABS is also prepared as a command-line interface (CLI). The CLI variant supports both Excel and JSON as input and outputs the result in JSON. A working Python environment is needed to install and run the NALABS CLI.

4.4 Progim

Progim will design Trace UML models and architectures to establish the structural framework for data tracking and simulation, ensuring alignment with industrial requirements. Furthermore, they will play a critical role in the validation phase by leveraging their deep domain knowledge to compare and verify the outputs of the developed applications, specifically conducting rigorous testing for Root Cause Analysis and Anomaly Detection to ensure the analytical results accurately reflect real-world industrial scenarios.

4.5 University of Szeged

The University of Szeged, Department of Software Engineering, will serve as a technology provider, leading the research and development of advanced anomaly detection methodologies grounded in both statistical modeling and machine learning. Their work builds on substantial practical experience in detecting and analyzing anomalies in real-world systems, including consistency verification of network packet streams, identification of zero-day cyberattacks, and monitoring of smart metering infrastructure. This expertise extends to detecting irregular electricity consumption patterns and fuel theft, leveraging both online (streaming) data and offline, retrospectively available datasets. In addition, the department will contribute cutting-edge expertise in applying large language models to formalize naturally expressed models, enabling the transformation of informal, human-readable descriptions into precise, machine-interpretable representations. This capability will further support systematic model comparison, consistency analysis, and validation, ensuring that heterogeneous models can be rigorously evaluated and aligned within a unified analytical framework.

4.6 HCL

HCL is a technology provider in MONA-LISA, mostly through two tools for the development of real-time software in C++:

- Code RealTime is available for IDEs that support Visual Studio Code extensions (e.g., Cursor, Eclipse Theia, DevOps Code, and of course, VS Code itself). Code RealTime is available in both commercial and community editions (free for non-commercial use).
- Model RealTime, available for the Eclipse IDE. Model RealTime is a commercial tool.

Both these tools share the same run-time library, and HCL will extend it with a new tracing capability. This enables capturing run-time traces from real-time applications developed with these tools. For details on how this works, see this page.

Captured traces can be visualized as sequence diagrams in Code RealTime, and a limited set of analysis capabilities will also be provided in that tool. For more advanced analysis and visualization, the traces can be translated by open source scripts, for example to the Google Trace Event format, which is the key tracing format used in MONA-LISA.

4.7 Alpata

Alpata will facilitate the integration of LTG-provided data into the middleware layer and subsequently leverage it to develop Machine Learning and Deep Learning models tailored for Root Cause Analysis and Anomaly Detection. They will establish the operational pipeline for these models in alignment with the MONA LISA Infrastructure, ensuring seamless data processing and interoperability. The insights and outputs generated by these analyses will serve as the foundational input for the visualization and interactive analytics components in WP6.

5. Summary of Analysis needs and available tools, technology and knowledge

This section summarizes the analysis needs from the use case providers and the tools, technologies and knowledge that will shape the solutions.

5.1 Summary of analysis needs in MONA LISA

Table 1 summarizes the analysis requirements in MONA LISA.

Table 1. Types of MONA LISA analysis

Use case provider	Type of analysis
Alstom	Automated root cause analysis of defects
Eli Alps	Automated root cause analysis of defects
Bilecik Demircelik	Automated root cause analysis of defects & predictive analysis
CNET	Predictive analysis of remaining useful life

5.2 Summary of available analysis tools, technology and knowledge in MONA LISA

Table 2 provides a summary of the available tools, technology and knowledge in MONA LISA.

Table 2. MONA LISA tools, technology and knowledge

Tools, technology and knowledge provider	Description
Evosoft	Final-layer visualization of processed data and analytics development
Lider	Data extraction and processing for analysis
Mälardalen University	Extensions of tools DRACONIS and NALABS for automated root cause analysis
Progim	Validation of automated root cause analysis and anomaly detection solutions
Uni. of Szeged	Anomaly detection and LLM-based data transformation
HCL	Capture and analysis of runtime traces from realtime applications developed in Code RealTime and Model RealTime tools
Alpata	Machine learning and deep learning models for automated root cause analysis and anomaly detection