GRAPHWISE
AI THRIVES ON WHOLE DATA

Vladimir Alexiev, PhD, PMP
Chief Data Architect

UNDERPIN
DATA SPACE FOR MANUFACTURING

Using LLM to Generate In-Depth Column Descriptors in the UNDERPIN Manufacturing/Maintenance Dataspace

Large Language Models to support semantic interoperability
Interoperable Europe, SEMIC and BDVA Workshop
10 April 2025

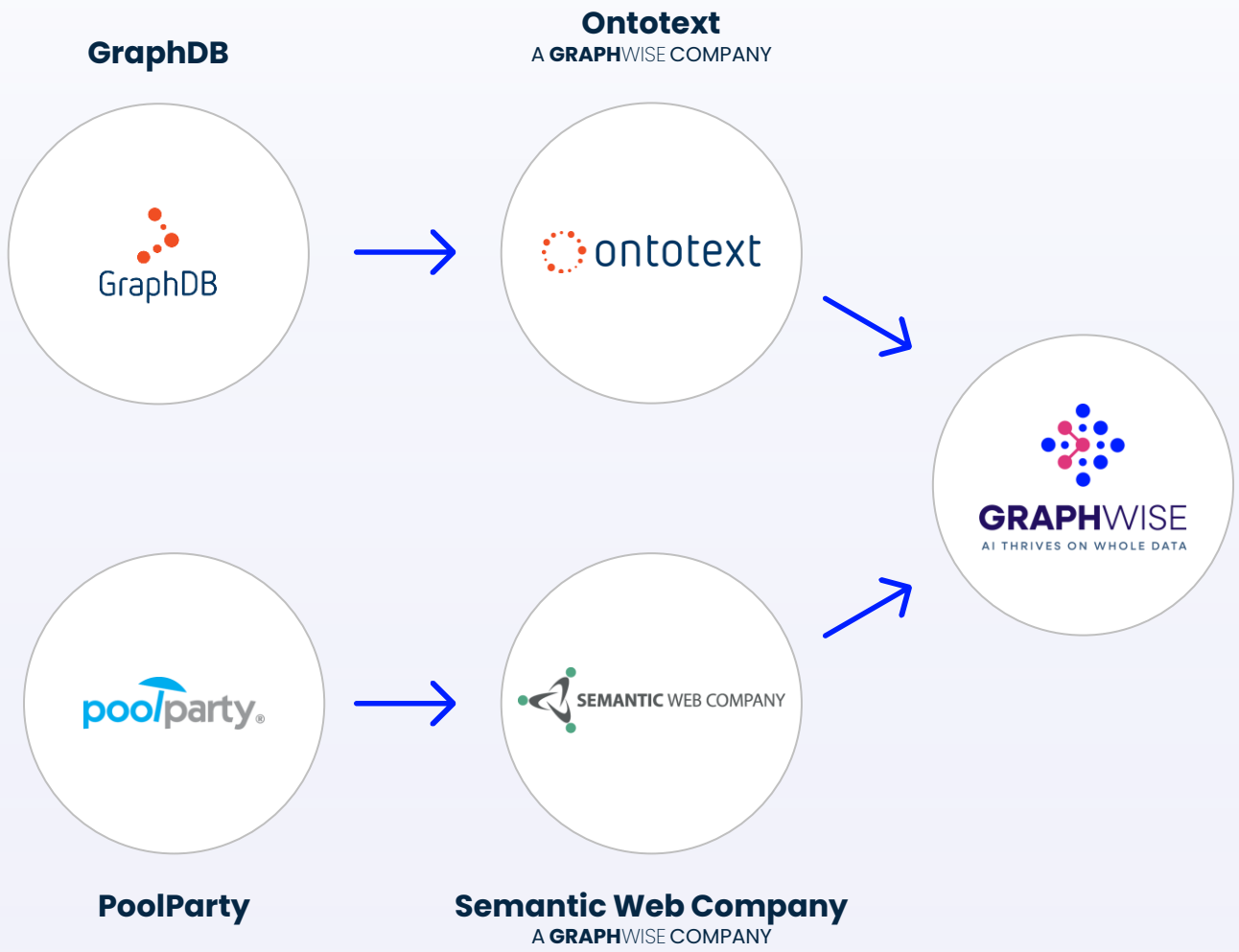
interoperable europe

SEMIC

One Stop Shop for GraphAI

GRAPHWISE
AI THRIVES ON WHOLE DATA

Graphwise has the platform and tools, the methodologies, and the people to deliver world-class end-to-end solutions to enterprises.



UNDERPIN Dataspace

GRAPHWISE
AI THRIVES ON WHOLE DATA

- Critical energy resources
  - Refineries, wind farms
- Tabular datasets
  - Sensor maintenance data
  - Predictive analytics
- Implemented **value-added services**
  - Semantic model that goes beyond DCAT
  - MinIO file store, Influx timeseries database
  - PoolParty Vocabulary Hub, GraphDB metadata store
  - Data vault, semantic search, GraphDB & Influx ingest, visualization dashboard
- Metadata facilitates
  - Rich discovery: semantic faceted search
  - Influx ingest: multiple CSVs to same timeseries
  - More flexible predictive analytics

UNDERPIN Standard Software

GRAPHWISE
AI THRIVES ON WHOLE DATA

- **Authority Portal**: Central point of the dataspace, with federated catalog
- **Dataspace Connectors**: one per participant
- **Keycloak**: Central authorization point used by the other software
- **Hashicorp Vault**: Repository for secrets and credentials for all other software
- **InfluxDB with Telegraf**: Time-series database, used for more powerful predictive analytics
- **File storage S3 MinIO**: File store (CSVs) for the dataspace. Open-source implementation of S3
- **File storage S3 API**: API for posting files to dataspace
- **GraphDB**: Stores semantic metadata (schemas and datasets) and vocabularies
- **Ontotext Refine**: RDFization of tabular metadata
- **PoolParty**: Vocabulary Hub: management of Underpin vocabularies and reference ontologies
- **Elasticsearch**: Full-text and semantic search, fed using the GraphDB Elastic connector
- **Chronograf**: Dashboard visualization (charts), comes with Influx
- **Kibana**: Dashboard visualization (charts), part of the Elastic stack
- **ArgoCD**: Kubernetes Controller for Continuous Delivery and Deployment

UNDERPIN Custom software

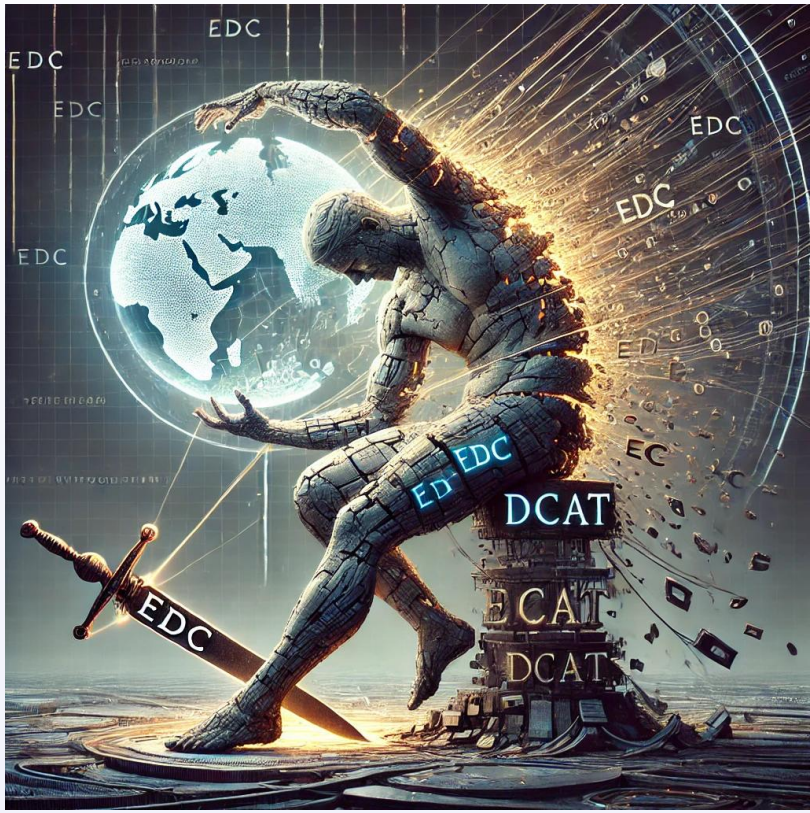
GRAPHWISE
AI THRIVES ON WHOLE DATA

- **GraphDB Metadata Ingest**: Interrogate Federated Catalog and ingest to GraphDB
- **Data Vault**: Central file storage (MinIO S3 as backend).
- **Influx Ingest**: Use metadata to Ingest CSV to Influx timeseries
- **Structure Definition (Schema) Editor**: allow user to create, edit or reuse a table schema using CSVW vocabulary.
- **Semantic Search**: Semantic dataset search, including facets, full-text, keywords. Uses Elasticsearch, fed by GraphDB Elastic connector
- **Semantic Model**: Specifies Underpin semantic metadata, generates RDFization scripts from declarative models
- **Visualization Dashboard**: Visualize input and result datasets from Influx
- **Blockchain with Smart Contracts**: Cryptographic non-repudiation of accepting a data contract and accessing a dataset. Store information about important events concerning data exchange
- **Refinery Predictive Analytics**: Predict anomalies of refinery compressors based on sensor data
- **Windfarm Predictive Analytics**: Predict anomalies of wind turbine generators based on sensor data

Semantic Interop in Dataspaces

GRAPHWISE
AI THRIVES ON WHOLE DATA

- Dataspace interop remains elusive
  - Vocabulary Hub, if used, is only a container for ontologies and vocabularies
  - EDC metadata payload not suitable as RDF (DCAT+ODRL)
  - EDC implementation and conformance test (TCK) don't use semantic approaches
  - V.Alexiev. Semantic Problems in Dataspaces. AIOTI Workshop on Semantic Interoperability for Digital Twins, 5-6 February 2025, Sophia Antipolis, France. paper, presentation
- UNDERPIN Metadata
  - Dataset description: "real" DCAT, PROV relations
  - Column descriptions: CSVW



An allegorical representation of DCAT as Atlas who underpins the sharing of datasets worldwide; but in a crippled state with missing parts, due to its diminished adoption by the Eclipse Dataspace Components, symbolized as a sword labeled "EDC".

Anyone can be a Picasso by using LLMs

UNDERPIN Semantic Search

GRAPHWISE
AI THRIVES ON WHOLE DATA

- Full-text, nested facets, auto-completion

```
- fieldName: text
  propertyChain: >-
    dct:identifier |
    dct:title |
    dct:publisher/schema:name |
    dct:temporal/dcat:startDate |
    dct:temporal/dcat:endDate |
    dct:conformsTo/dct:title |
    dct:conformsTo/csvw:column/dct:title |
    dcat:keyword
  datatype: native:text
- fieldName: text_completion
  propertyChain: "@text"
  datatype: native:completion
  analyze: true
  nativeSettings:
    preserve_separators: true
    preserve_position_increments: true
    max_input_length: 50
- fieldName: column
  propertyChain: dct:conformsTo/csvw:column
  array: true
  datatype: native:nested
  objectFields:
  - fieldName: features
    propertyChain: sosa:hasFeatureOfInterest/skos:prefLabel
    array: true
    analyzed: false
  - fieldName: qualifiers
    propertyChain: un:qualifier/skos:prefLabel
    array: true
    analyzed: false
  - fieldName: quantity
    propertyChain: qudt:hasQuantityKind/skos:prefLabel
    analyzed: false
  - fieldName: unit
    propertyChain: qudt:hasUnit/skos:prefLabel
    analyzed: false
```

+ More

KEYWORDS

- Temperature 32
- degree Celsius 32
- Average 31
- Bearing 31
- Revolution per Minute 31

+ More

TYPES

- CSV dataset 28
- Input dataset 26
- Result dataset 8
- Influx dataset 6

PUBLISHER

- Motor Oil Renewable Energy (MORE) 14
- Motor Oil (MOH) 13
- SpaceHellas 6
- AIT Austrian Institute Of Technology 1

TAG

- WF1-WTG01 4
- WF1-WTG05 3
- WF1-WTG02 1
- WF1-WTG03 1
- WF1-WTG04 1

+ More

START DATE

- October 2019 (2)

ID: refinery-compressor-result-anomaly-2022-12.csv
Keywords: Anomaly, Compressor, CSV dataset, Predicted, Refinery, Result dataset
Publisher: SpaceHellas

Refinery Compressor Sensor Data For 2022-06 As Csv

ID: refinery-compressor-2022-06.csv
Keywords: ampere, angular velocity, Average, Axial, Bearing, Blow Off, Compressed Air, Compressor, Control Oil, CSV dataset, degree Celsius, Differential, Discharge, Displacement, distance, electric current, Gearbox, Inlet, Input dataset, Journal Bearing, Journal Pads, K-2201/KT-2201, K-3201 A/KM-3201 A, K-3301 B/KT-3301 B, K-5701, K-7502/ST-7501, Kilogram Force per Square Centimetre, Kilogram per Hour, Lube Oil, Main Pump, Mass Flow Rate, Maximum, Micrometre, Millimetre, Minimum, Motor, Outlet, pressure, Primary Vent, Radial, Refinery, Revolution per Minute, Seal Oil, Shaft, Stator, Steam Turbine, Temperature, Thrust Bearing, Thrust Pads, Turbine, Vibration
Publisher: Motor Oil (MOH)

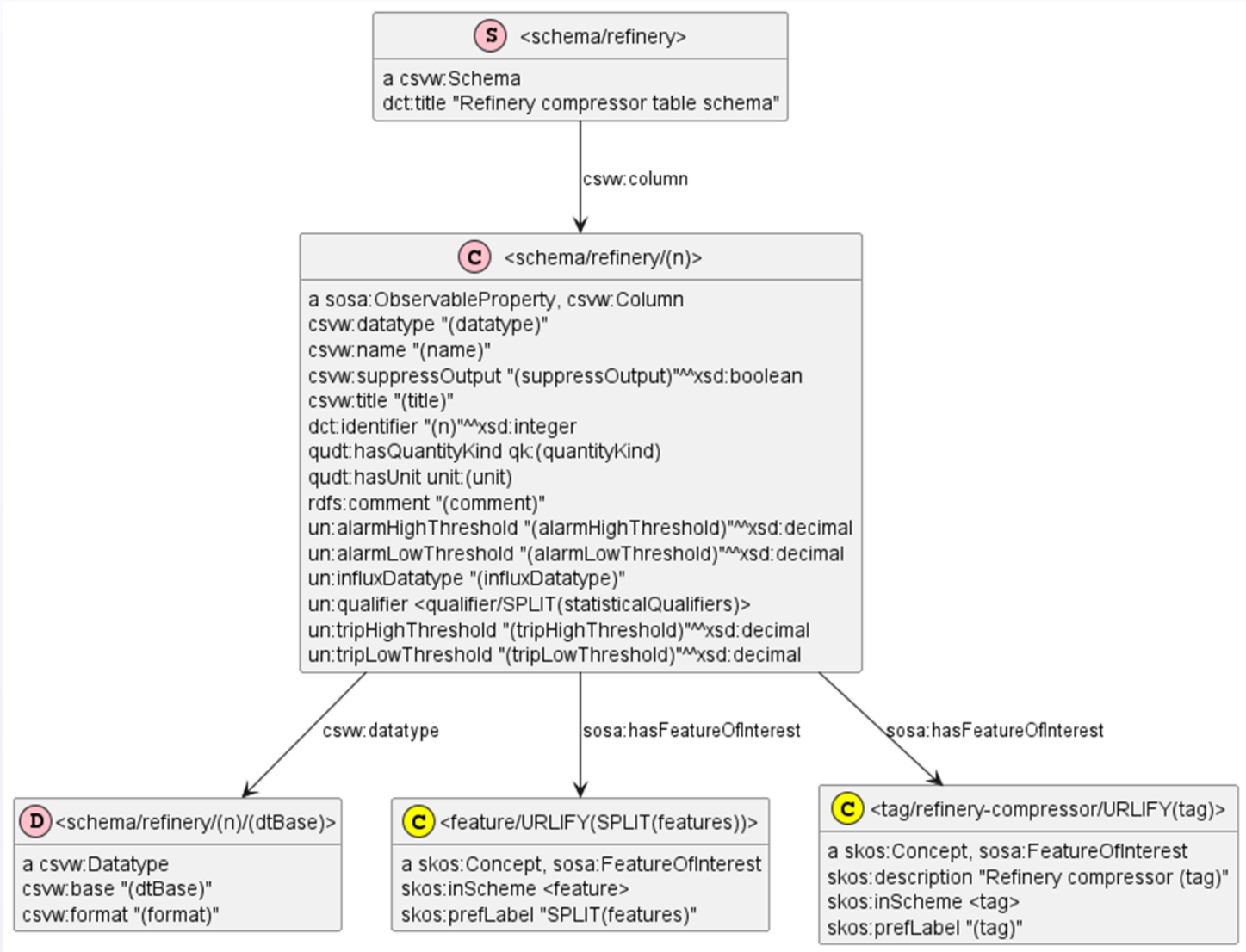
Refinery Compressor Sensor Data For 2022 As Influx

ID: refinery-compressor-2022-influx
Keywords: ampere, angular velocity, Average, Axial, Bearing, Blow Off, Compressed Air, Compressor, Control Oil, degree Celsius, Differential, Discharge, Displacement, distance, electric current, Gearbox, Influx dataset, Inlet, Input dataset, Journal Bearing, Journal Pads, K-2201/KT-2201, K-3201 A/KM-3201 A, K-3301 B/KT-3301 B, K-5701, K-7502/ST-7501, Kilogram Force per Square Centimetre, Kilogram per Hour, Lube Oil, Main Pump, Mass Flow Rate, Maximum, Micrometre, Millimetre, Minimum, Motor, Outlet, pressure, Primary Vent, Radial, Refinery, Revolution per Minute, Seal Oil, Shaft, Stator, Steam Turbine, Temperature, Thrust Bearing, Thrust Pads, Turbine, Vibration

Table Schema model

GRAPHWISE
AI THRIVES ON WHOLE DATA

- Reused ontologies
  - CSVW, DCT, RDFS QUDT, SKOS
- Column characteristics:
  - **Features**: parts, components, consumables, ambient (air, wind)
  - **Tags**: specific components
  - **Qualifiers** (statistical etc): avg, min, max, stdDev, relative, absolute, differential, total...
  - **Quantity Kind**: Temperature, Pressure, Displacement...
  - **Unit**: DEG_C, KiloGM_F-PER- CentiM2, MilliM...
  - **Thresholds**: low/high, alarm/trip
  - csvw:**datatype**, un:influxDatatype
  - csvw:**suppressOutput**: don't ingest in Influx



Raw vs semantically explicated columns

GRAPHWISE
AI THRIVES ON WHOLE DATA

raw name	fixed name	features	qualifiers	kind	unit
WTG01_Generator Bearing2 Temp. Avg. (1)	Generator Bearing2 Temperature Average	Generator, Bearing	Average	Temperature	DEG_C
WTG01_Generator RPM Max. (2)	Generator RPM Maximum	Generator	Maximum	Angular Velocity	REV-PER-MIN
WTG01_Ambient WindSpeed Max. (18)	Ambient Wind Speed Maximum	Wind	Maximum	LinearVelocity	M-PER-SEC
WTG01_Grid ok hours (23)	Grid ok hours	Grid	Runtime	Time	HR
WTG01_Total Active power (30)	Total Active power		Total	ActivePower	W
WTG01_Production LatestAverage Total Active Power Avg. (82)	Production Latest Total Active Power Average	Production	Average	ActivePower	W
WTG01_Total reactive power (83)	Total reactive power		Total	ReactivePower	V-A_Reactive
WTG01_Production LatestAverage Total Reactive Power Avg. (84)	Production Latest Total Reactive Power Average	Production	Average, Latest, Total	ReactivePower	V-A_Reactive
75TI834.pv	COMPRESSOR BCL 509/A JOURNAL BEARING TEMPERATURE	Compressor, Journal Bearing	Maximum	Temperature	DEG_C
75XI823BX.pv	COMPRESSOR BCL 509/A JOURNAL BEARING RADIAL VIBRATION	Compressor, Journal Bearing, Vibration, Radial	Maximum	Distance	MicroM
75ZI800BB.pv	STEAM TURBINE NG 40/32 SHAFT AXIAL DISPLACEMENT	Steam Turbine, Shaft, Axial	Average	Displacement	MilliM
75PI870.pv	CONTROL OIL HEADER PRESSURE	Control Oil	Minimum	Pressure	KiloGM_F-PER-CentiM2

Explicating Columns with LLM

GRAPHWISE
AI THRIVES ON WHOLE DATA

- I don't know what is "D.E. BEARING TEMP AVG"
- But an LLM not only knows:
  - With simple prompting it can create a proper concept:
  - Characteristics, definition, scope notes in regard to monitoring and significance
- Example:
  - csvw:title "D.E. BEARING TEMP AVG"
  - csvw:name "Drive-End Bearing Temperature Average"
  - rdfs:comment "Temperature of the bearing located on the drive-end (D.E.) of a rotating machine, such as a motor, pump, or turbine";
  - skos:scopeNote "Monitoring: The drive-end is where the motor or shaft connects to the load; excessive heat in this bearing can indicate misalignment, poor lubrication, or wear.";
  - skos:scopeNote "Significance: High bearing temperature can lead to bearing failure, resulting in machine downtime or damage.".
  - sosa:hasFeatureOfInterest <feature/bearing>, <feature/drive-end>
  - un:qualifier <qualifier/average>
  - qudt:hasQuantityKind qk:Temperature
  - qudt:hasUnit unit:DEG_C
  - csvw:datatype "double"

Acknowledgement

The UNDERPIN Pan-European data space for holistic asset management in critical industries has received funding from the Digital Europe Programme under grant agreement No 101123179.

https://www.graphwise.ai
https://www.ontotext.com | https://www.poolparty.biz
Vladimir Alexiev - vladimir.alexiev@graphwise.ai