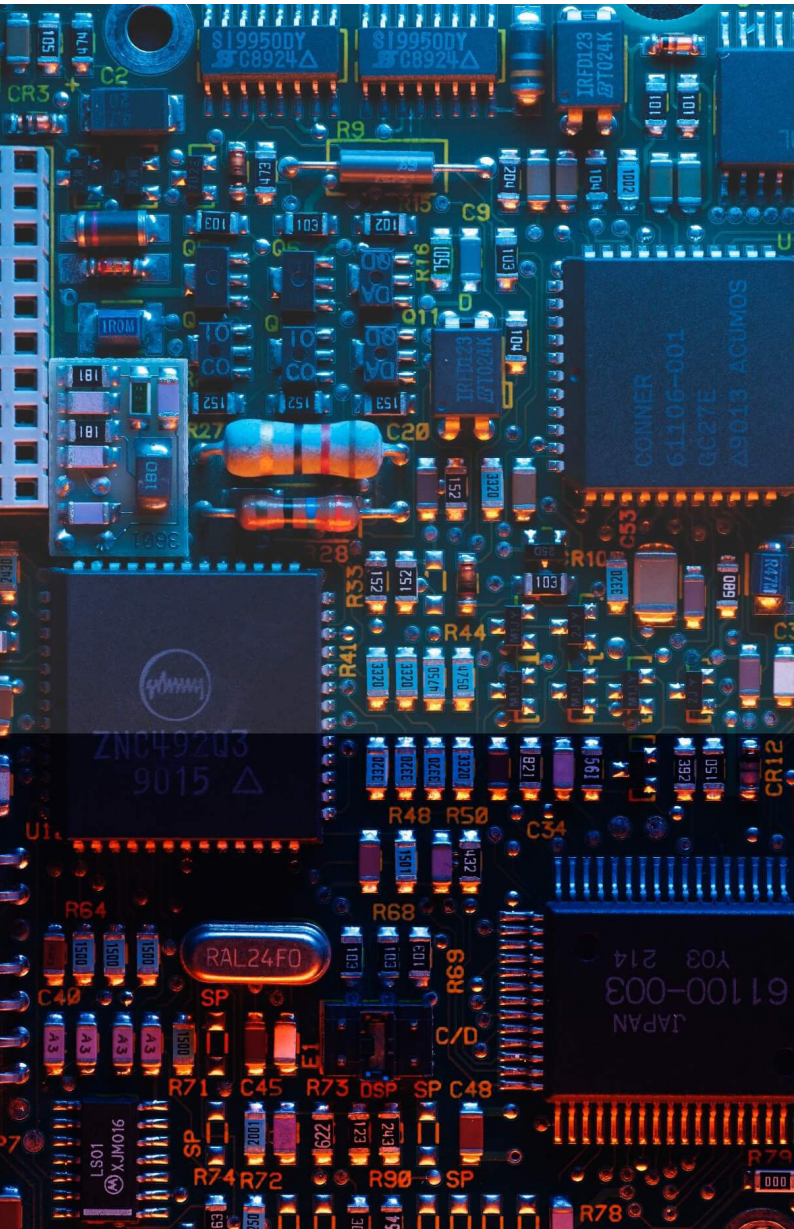


Operational Excellence

## Smart Innovators: Industrial Data Management Solutions

By Joe Lamming  
With Malavika Tohani

February 2023



# Smart Innovators: Industrial Data Management Solutions

By Joe Lamming  
With Malavika Tohani

February 2023

Data represents reality, but is not reality itself. It consists of symbols and numbers used to convey information, but is not the information itself. As a human-made construct, data is often manipulated and transformed to serve a specific purpose, which can compromise its authenticity or accuracy. Without proper processes in place to manage access, track transformations and provide context to human operators, the unstoppable universal truth of increasing entropy will begin to introduce bias and error into the data, further distorting its representation of reality. To aid customers in navigating the complex and rapidly evolving field of industrial data management, this report examines the capabilities of 20 vendors offering industry-specific DataOps, data hub and data analytics solutions. In addition, it highlights the data management priorities of industrial firms, considering what software capabilities exist to address real-world needs and how these can be combined to provide a best-fit solution. C-Suite executives, enterprise data science teams and operations managers should consider the solutions discussed in this report to support industrial data strategy.

## Table of contents

<b>Data Is Most Valuable When In-Formation, And Information Is An Unnatural State</b>	3
Industrial Operations Are More Data-Driven Than Ever	
Data Management Is An Increasing Priority For Industrial Firms	
<b>Introducing The Industrial Data Management Solutions Market</b>	5
Solutions For Industrial Data Management Comprise Three Segments	
Industrial Data Management Solutions Enable Scalable, Secure, Inter-Asset Data Flow	
Industrial Data Management Solutions Have 10 Key Capabilities	
Customers Must Develop Vision And Strategy Before Beginning Their Data Management Journey	
Vendors Should Expand Data Acquisition Towards Varied, External Sources	

## Table of figures

<b>Figure 1.</b> Data Management Priorities For 2023	4
<b>Figure 2.</b> Industry Coverage For Industrial Data Management Solutions Providers	6
<b>Figure 3.</b> The Landscape Of Data Hubs, DataOps Platforms And Data Analytics Software	7
<b>Figure 4.</b> Industrial DataOps Platforms Orchestrate People, Assets, Processes And Technology	8
<b>Figure 5.</b> Capabilities Criteria For Industrial Data Management Solutions	11
<b>Figure 6.</b> Vendor Capabilities: Industrial Data Management Solutions	13

## Organizations mentioned

3d Signals, Accenture, AccuRounds, Allen-Bradley, Arkema, AspenTech, Augury, AVEVA, BASF, C3 AI, Celanese, Chevron Phillips Chemical, Cognite, Colonial Pipeline, DataProphet, deepset, Element Analytics, Etalab, Google, HENN, HighByte, Hitachi Vantara, Honeywell, IBM, inmation, MachineMetrics, Mars, Microsoft, Morgan Stanley, OpenAI, Pacific Gas & Electric (PG&E), Palantir, PTC, Radial SG, Seeq, Siemens, Skkynet, SLB (formerly Schlumberger), Software AG, SparkCognition, SymphonyAI Industrial, Teel Plastics, TEPSYS, TrendMiner, Uptake, US Securities and Exchange Commission (SEC), Yokogawa RAP, YuzeData.



# Data Is Most Valuable When In-Formation, And Information Is An Unnatural State

Data – the plural of datum – from the Latin word for ‘thing given’, has seen its meaning evolve from a reference to a cluster of observations, towards a synonym for ‘information’. While originally a plural noun, the word ‘data’ is so frequently used in the context of uncountably large quantities of observations that it has become a mass noun (‘this data’ vs ‘these data’). ‘Information’, on the other hand, is more abstract, referring to observable patterns interpreted to have meaning. Without interpretation and without understanding, data loses all meaning. For example, in the context of industrial data, a temperature measurement from a sensor will simply present itself as a number. If, however, such measurements are aggregated with metadata (data that gives information about other data), contextualized in an asset hierarchy, and annotated with notes from operations and maintenance (O&M) personnel, these individual data streams are converted into meaningful observations and information, which can be acted upon.

Sectors primarily focused on analysing numbers were among the first to utilize algorithmic decision-making in day-to-day operations. The life insurance industry began using computers in the 1950s to optimize premiums by helping discover trends in large data sets, while financial firms introduced market-driven computational feedback loops to improve arbitrage trading in the 1960s. In the 21st century, large technology firms have pioneered the use of machine learning (ML) to find patterns in billions of actions performed by hundreds of millions of web users, using this for targeted advertising. Driven by increasing costs, new technologies, labour shortages and supply chain disruption, asset-heavy industries are increasingly seeing the value in performing data analytics to achieve safe, efficient and sustainable operations.

## Industrial Operations Are More Data-Driven Than Ever

In the industrial world, feedback loops as part of advanced control systems and asset management software have driven the rise of automation in nearly every sector – helping improve overall equipment effectiveness (OEE) through 24/7 operation and optimized maintenance processes in the context of labour shortages (see [Verdantix Global Corporate Survey 2022: Operational Excellence Budgets, Priorities & Tech Preferences](#)). Over the next 12 months, industrial data management practices will be turbo-charged by a range of factors, as:

- **Industrial facilities generate more data than ever before.**

Manufacturing facilities are generating more and more operational technology (OT) data, enabled by inexpensive sensors, multiple control systems and human-machine interfaces. Morgan Stanley estimated in 2017 that manufacturing firms generated two exabytes of data annually, but discarded 99% of it. Six years later, industrial facilities are more connected, with access to cheaper storage and more sophisticated data management tools – offering operators a platform from which to glean game-changing insights. DataProphet worked with an automotive alloy wheel original equipment manufacturer (OEM) to optimize production, by using a data collection system to monitor a variety of production parameters. PRESCRIBE, DataProphet’s solution, delivered real-time adjustments to control systems to reduce scrap by 29%.

- **Out-of-the-box AI and ML tools boost data discovery.**

AI, sometimes used synonymously with ML and deep learning, is today commonly developed by ‘training’ weights and biases in an artificial neural network. Throughout 2022, the startling capabilities of generative natural language processing (NLP) models such as OpenAI’s GPT-3.5 and Google’s LaMDA dominated global headlines. In the industrial context, such NLP models offer firms semantic search capabilities for asset documentation and for analysing trends in written inspection reports. Other ML techniques focused on optical character recognition (OCR) and symbol identification have witnessed widespread industrial use in enriching piping and instrumentation diagrams (P&IDs) with searchable text and in automatically creating associations between documents, images and asset maintenance schedules. Cognite and Radial SG offer P&ID enrichment to provide industrial firms with contextualized data in a single graphical user interface (GUI).



- **Firms acquire data with greater volume, velocity, variety and visibility.**

Increasing the number of sensors – and the rate at which these sources are polled for new data – will result in a more granular representation of assets within a facility. However, a mountain of raw sensor feeds is useless without variety. Industrial firms looking to maximize the utility of OT data must combine this with real-time and historical images, audio, video, documents and logs from wearable technology. Making all of this visible and understandable across an enterprise will help O&M teams plan further ahead, assist data scientists in selecting the most relevant AI/ML training data sets, and inform C-Level executives on the overall health of a business (see [Verdantix Why Industrial Firms Need DataOps Platforms For Asset Management Digitization](#)).

- **High-quality OT data drives successful industrial AI deployments.**

Model training for applications such as predictive maintenance is performed with vast quantities of data, with real-world inputs – a group of sensor readings – producing an actionable output such as ‘pass’ or ‘fail’. Real-world performance of such application-specific models is highly sensitive both during training and while deployed in production, and with less data from which to train models, the importance of eliminating signal noise, errors and sensor drift, and identifying and eliminating irrelevant features, becomes critical. Industrial AI and ML models designed to detect an upcoming pump failure, for example, will benefit from advances in data management and data cleaning. As data management best practices are enforced through software tools, well-planned digitization projects and AI-driven techniques, firms will improve the discoverability, variety and quality of industrial data – and make new data sets available to train next-generation tech.

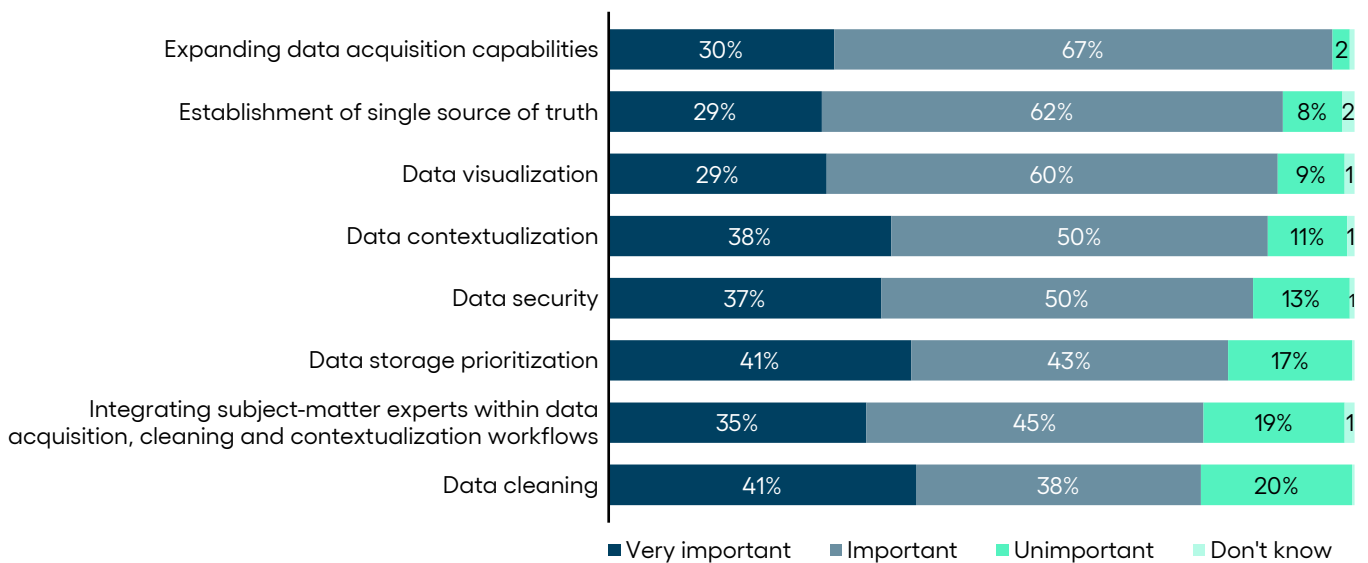
## Data Management Is An Increasing Priority For Industrial Firms

Today, industrial firms have at their disposal the ability to quickly deploy thousands of sensors across physical assets, and the software to make sense of it all. The 2022 Verdantix global corporate survey explored the data management activities of 301 senior executives at industrial firms (see **Figure 1**) (see [Verdantix Global Corporate Survey 2022: Operational Excellence Budgets, Priorities & Tech Preferences](#)). We heard that in 2023 firms will:

Figure 1

### Data Management Priorities For 2023

**How important are the following data management activities for your firm over the next 12 months?**



Note: Data labels are rounded to zero decimal places; percentages less than 3% are written as numbers.

Source: Verdantix Global Corporate Survey 2022: Operational Excellence Budgets, Priorities & Tech Preferences

N= 301



- **Expand data acquisition capabilities.**

Deploying more sensors, cameras and connected worker solutions will be key to expanding the volume, velocity and variety of OT data in 2023. Ninety-seven per cent of our survey respondents described the expansion of data acquisition capabilities as ‘important’ or ‘very important’ activities over the next 12 months. Wisconsin-based Teel Plastics deployed PTC’s Kepware IoT Gateway to connect to machine programmable logic controllers (PLCs) by Allen-Bradley and Siemens. The solution automates much of the data acquisition, machine set-up and calibration processes and allows operators to combine additional data such as humidity and parts wear, to predict asset failure as well as optimize production.

- **Establish an auditable and fast-synchronizing single source of truth (SSOT).**

Ninety-one per cent of respondents to our survey said that establishing a single source of truth was an ‘important’ or ‘very important’ activity for 2023, while 84% mentioned data storage prioritization as ‘important’ or ‘very important’. SSOT architectures involve the creation of a centrally managed catalogue of references to original data storage locations, enabling data consistency, quality and efficiency by actively monitoring changes and enforcing synchronization, deduplication and integration across disparate enterprise storage systems. Skkynet offers SSOT functionality through its DataHub Transfer Protocol, which prioritizes low-latency, secure, bidirectional OT data synchronization and visualization across global sites.

- **Curate data from disparate sources with software-based contextualization.**

A high-resolution satellite image of Europe contains a vast array of data; however, without labels, graphical overlays of transportation networks and interactive sub-menus, its utility to a supply chain decision-maker is low. Without context, operators will struggle to transform data into information and thus into knowledge. Eighty-eight per cent of survey respondents mentioned data contextualization as a ‘very important’ or ‘important’ activity for their firms in 2023. Palantir worked with Pacific Gas & Electric (PG&E) to centralize, curate and make sense of nearly 10 billion data points collected from more than 16 million customers, to optimize its grid operations. Palantir’s Foundry also provides predictions on fire risks across 25,000 miles of distribution infrastructure, guiding PG&E to switch off equipment only when necessary.

- **Integrate data from varied sources with help from subject-matter experts.**

As enterprise asset management (EAM) becomes more digitized, the sheer variety and volume of data will become unmanageable for any single stakeholder. When asked about data management plans for 2023, 80% of survey respondents said that integrating subject-matter experts (SMEs) into data acquisition, cleaning and contextualization workflows would be a ‘very important’ or ‘important’ activity. Without clear annotations and context for OT data provided by those most familiar with it, the development and deployment of new analytical techniques will be slowed by clarification requests via emails, messages and phone-calls.

## Introducing The Industrial Data Management Solutions Market

Data is rapidly becoming fundamental to the modern industrial world, but spreadsheets, emails and phone-calls alone are unable to scale with firms’ digital ambitions. EAM systems help streamline asset registry, resource and work order management (see [Verdantix Green Quadrant: Enterprise Asset Management Software 2022](#)), while enterprise resource planning (ERP) and manufacturing execution systems (MES) provide visibility to managers on activities such as scheduling, inventory, production and quality. However, the differences between how these systems operate and who uses them creates silos of data within facilities and across global enterprises. Bringing together the priority activities – data acquisition, SSOT, context and SME integration – requires a more holistic solution.

To gain an in-depth understanding of the industrial data management solutions market, Verdantix evaluated 20 industrial-focused data management software providers. The following 17 vendors participated actively in the research: AspenTech, AVEVA, C3 AI, Cognite, DataProphet, Element Analytics, HighByte, Hitachi Vantara, Honeywell, MachineMetrics, Radial SG, Seeq, Skkynet, SymphonyAI Industrial, Software AG, Uptake and YuzeData.



Vendors who were contacted but did not respond, or missed deadlines, were IBM, Palantir and PTC. Information provided during briefings and through publicly available case studies shows how vendors included in the analysis cover a variety of industries (see **Figure 2**).

**Figure 2**  
Industry Coverage For Industrial Data Management Solutions Providers

Firm	Software Name	Oil & Gas	Process Manufacturing	Discrete Manufacturing	Infrastructure & Transport	Power Utility	Water & Wastewater Utility	Mining & Metals
AspenTech	Inmation	●	●	◐	○	◐	○	○
AVEVA	Data Hub, PI System	●	●	◐	◐	◐	◐	◐
C3 AI	Data Vision	●	◐	◐	○	◐	○	○
Cognite	Data Fusion	●	◐	◐	○	◐	○	○
DataProphet	CONNECT	○	◐	●	○	○	○	◐
Element Analytics	Unify	●	●	○	○	◐	○	◐
HighByte	Intelligence Hub	○	◐	●	○	○	○	○
Hitachi Vantara	Lumada Industrial DataOps	◐	◐	●	◐	●	◐	◐
Honeywell	Forge, Matrikon	●	◐	○	○	○	○	◐
IBM	Watson Knowledge Catalog, Maximo Monitor	●	○	◐	○	●	○	◐
MachineMetrics	Industrial IoT Platform	○	○	●	○	○	○	○
Palantir	Foundry	●	◐	◐	○	●	○	○
PTC	Kepware, ThingWorx	◐	●	●	○	○	○	◐
Radial SG	Viewport	●	●	○	○	◐	○	○
Seeq	Workbench, Data Lab, Cortex	●	●	○	○	◐	◐	◐
Skkynt	Cogent DataHub, SkkyHub	●	◐	◐	○	●	●	◐
SymphonyAI Industrial	EurekaAI Industrial Platform	◐	●	◐	○	◐	○	◐
TrendMiner	Analyze, Monitor, Contextualize	◐	●	○	○	○	◐	○
Uptake	Fusion	●	◐	○	●	◐	○	◐
YuzeData	YuzeData Platform	●	◐	◐	○	◐	○	◐

Not a focus	○
Some customer base	◐
Focus (with the majority of customer base)	●

Note: Industry coverage based on case studies and information provided during briefings, where available.  
Source: Verdantix analysis



# Solutions For Industrial Data Management Comprise Three Segments

The industrial data management space is filled with solutions offering varying capabilities and multiple terminologies, making it difficult for buyers to select one that matches their needs. Verdantix research finds that solutions for industrial data management fall into three distinct categories (see **Figure 3**). These are:

- Data hubs to quickly aggregate data from disparate sources.**  
 As point-to-point solutions are not scalable, adopting a hub-and-spoke data model – where disparate data-generating assets and systems connect to a centralized data broker – is more beneficial, as enterprise IT teams need only maintain a single connector for each data generator or consumer. Data hubs automatically enable interoperability between systems and provide a strong foundation on which to develop additional data transformation processes. Market-leading data hubs also enable SSOT, data storage management, lineage and data visualization. HighByte’s Intelligence Hub offers manufacturing firms a strong platform for configuring data ingestion pipelines with common machine protocols, and for viewing metadata, along with the ability to combine data sets into payloads for upload to cloud storage. Meanwhile, in February 2022 AVEVA launched its software-as-a-service (SaaS), cloud-based Data Hub, to integrate with its PI System, for customers to remotely view OT data distributed across global sites.
- DataOps platforms to maintain quality, context and discoverability.**  
 Industrial DataOps is a condensation of ‘Data’ and ‘Operations’, where the practice of aligning business operational goals with data orchestrates people, assets, processes and technology to receive and clean a wide variety of data streams, place these in context and deliver them as a self-explaining, digestible and trustworthy package to data stakeholders (see **Figure 4**). Leading DataOps platform vendors offer data hub functionality with pre-built, strong data quality and SSOT storage management, tried-and-tested cyber security, self-service contextualization and collaboration features. AVEVA, Cognite and Hitachi Vantara provide comprehensive industrial DataOps platforms for firms in manufacturing, oil and gas and power utilities.

Figure 3  
The Landscape Of Data Hubs, DataOps Platforms And Data Analytics Software

	Direct Industrial Data Acquisition	Indirect Industrial Data Acquisition	Single Source Of Truth	Data Storage Management	Data Cleaning, Lineage And Feature Extraction	Data Access Management And Security	Contextualization	External Data Sharing And Benchmarking	Data Analytics	Data Visualization
Data Hubs	●	●	◐	◐	◐	○	○	○	○	◐
DataOps Platforms	◐	●	●	◐	●	●	●	◐	◐	◐
Data Analytics	◐	●	○	○	◐	○	◐	◐	●	●

Won't have	○
Nice to have	◐
Could have	◑
Should have	◒
Must have	●

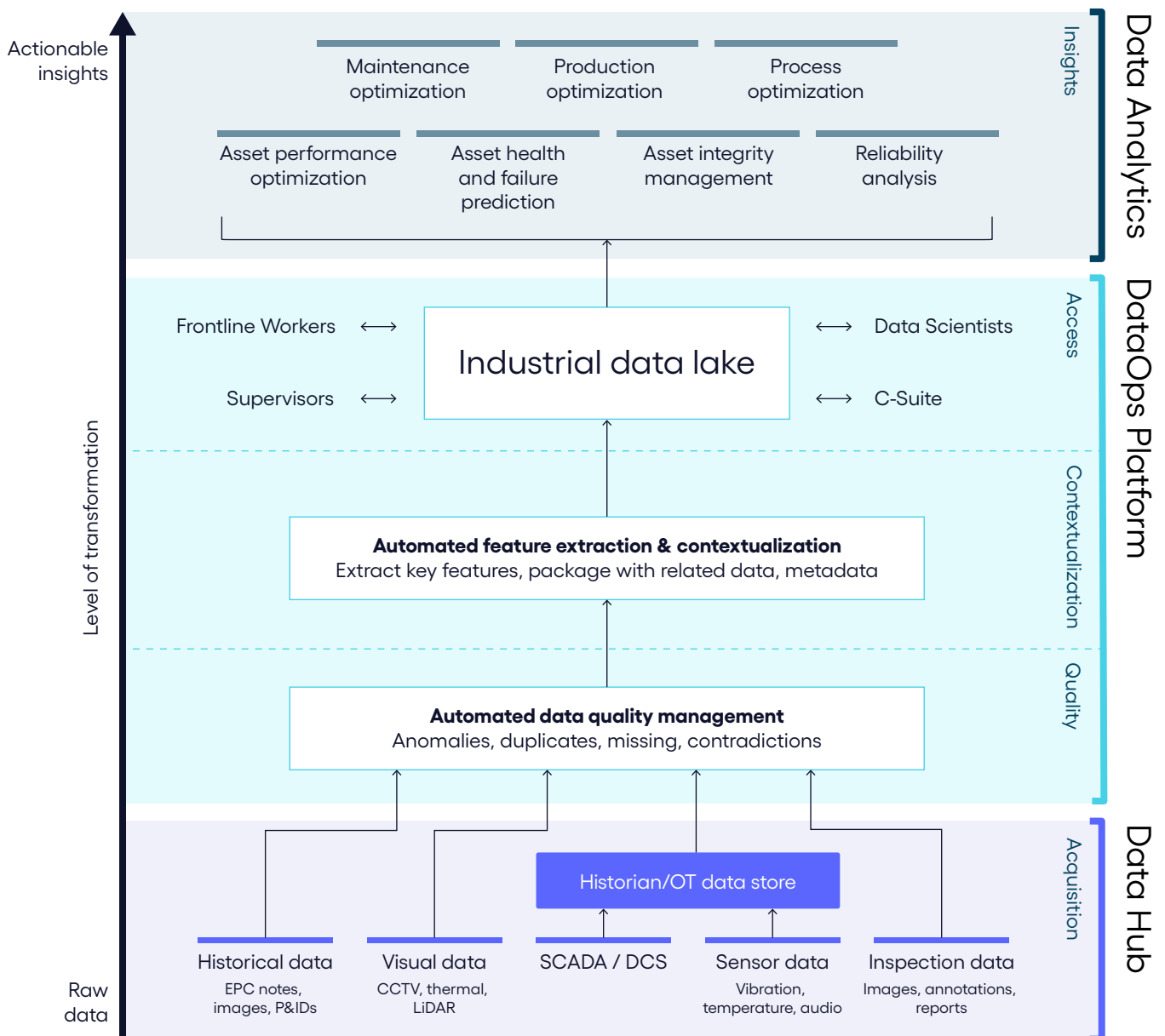
Source: Verdantix analysis



- **Data analytics to deliver insights through data processing and visualization.**

Industrial analytics – the systematic computational analysis of OT data for the purpose of generating actionable insights – are at the top of the data management stack. While some DataOps platform providers offer a wide range of ready-made analytics, the broad applicability of such data management software means that third-party analytics tools are still critical to delivering innovative or niche data processing and visualization. The rise of AI-focused analytics specialists such as 3d Signals, Augury and SparkCognition provides a plethora of pre-built, application-specific data quality management and feature extraction tools to allow plant managers and data scientists to quickly iterate and develop new insights.

Figure 4  
Industrial DataOps Platforms Orchestrate People, Assets, Processes And Technology



Source: Derivation of Verdantix Strategic Focus: Why Industrial Firms Need DataOps Platforms For Asset Management Digitization





# Industrial Data Management Solutions Enable Scalable, Secure, Inter-Asset Data Flow

Data hubs provide the integration layer between the data generators and the DataOps platform, while analytics perform the transformation and distillation of large data sets to deliver the human-readable and actionable insights. DataOps platforms are therefore at the heart of industrial data management. To meet customer requirements, industrial data management solutions providers must offer tools relevant to their unique data management challenges – be these data acquisition, cleaning, governance, contextualization, cyber security, benchmarking, analytics or visualization. Industrial data management solutions add value to industrial asset management by:

- **Ensuring data from connected workers, assets and facilities is scalable.**

Point-to-point solutions are intrinsically hindered by the need to individually configure and continuously maintain all data generators and data processors. By connecting data generators and processors to a common interface layer provided by a data hub or DataOps platform, the monitoring and maintenance of connections is largely automated – and new sensors can be added and managed from within a single platform. With the ability to remotely manage data acquisition across global sites, digitization projects become less reliant on duplicated IT teams and the scaling of deployments is made possible. In October 2022, AspenTech acquired industrial information management software vendor Inmation, as part of its expansion into real-time, operations-focused data management. Chemicals manufacturer BASF selected the firm's software to scale real-time connections and dashboards across hundreds of industrial and enterprise data sources and to replicate implementations across global sites.

- **Visualizing large volumes of varied OT data within a single pane of glass.**

Switching between software applications hinders the development of novel solutions by introducing friction to data interoperability. Data management can implement an enterprise-wide data discoverability and transformation platform, on which engineers, managers, data scientists and C-Level executives can upload, process and view data in context. Images, reports and P&IDs can be decorated with notes, sketches and references to time-series data, temperature readings and event logs. Cognite announced a partnership in September 2022 with SLB (formerly known as Schlumberger), to develop the ability of the oilfield services provider to integrate data from reservoirs, wells and facilities within Cognite's single-pane-of-glass Data Fusion platform. YuzeData, meanwhile, offers pre-built application-specific data management workflows, called YuzeCases, to integrate data from relevant sources within a single interactive user interface (UI).

- **Bringing the SME and data scientist together.**

Cross-functional collaboration within a common data ingestion, transformation, contextualization and visualization environment enables those most familiar with the physical phenomena to decorate the industrial data. DataOps platforms abstract the IT-focused difficulties involved with data discovery through automated enforcement of data management best practices and by providing more intuitive human-machine interfaces for both SMEs and data scientists to focus on and leverage their specific skill sets. Seeq worked with Chevron Phillips Chemical to reduce the complex array of Excel spreadsheets used in the firm's day-to-day operations. It leveraged data acquisition tools to connect directly to enterprise systems and to facilitate collaboration between SMEs and data scientists through self-service analytics – creating a culture of data-driven decision-making.



# Industrial Data Management Solutions Have 10 Key Capabilities

Industrial asset management sees strong value propositions offered by data management solutions such as data hubs, DataOps platforms and data analytics software, including scalable connectivity, strong visualizations and better data-driven enterprise collaboration. Verdantix carried out a review into the market for industrial data management solutions, exploring the pain points highlighted by our global survey, as well as trends in operational excellence technology (see [Verdantix Market Insight: 10 Predictions For Operational Excellence Technology In 2023](#)). We identified ten capabilities, spanning industrial data hubs, DataOps platforms and data analytics. Verdantix recommends that potential buyers assess vendors based on their ability to offer (see **Figure 5** and **Figure 6**):

## 1. Direct industrial data acquisition.

AVEVA PI System directly acquires data from assets and processes for pharma, oil, gas, power and mining firms. MachineMetrics aggregates and analyses data across its customer base, providing asset failure prediction and production optimization benchmarks.

## 2. Indirect industrial data acquisition.

Data hubs, DataOps platforms and data analytics solutions connect to historians and other business and IT systems, extracting near-real-time and historical data. Meanwhile, AspenTech, C3 AI, Cognite, Element Analytics, Hitachi Vantara, Honeywell and Uptake acquire similar data for use further up the stack. The majority of data management solution providers have built strong capabilities for indirect data acquisition, as this forms the foundation for strong analytics. SymphonyAI Industrial offers tools for pulling data from historians and EAM and ERP systems, to drive predictive maintenance and digital twins.

## 3. A distributed, consistent industrial data lake enabled by an SSOT.

Industrial data can be stored on-premise, or in centralized enterprise data centres, or distributed across multi-tenant cloud infrastructure. Data management solutions rely on the ability to gather and contextualize a wide variety of data from disparate locations – but gathering such data for specific tasks can result in duplicated data across an enterprise. Vendors with strong SSOT capabilities use robust, regular synchronization techniques and data conflict mitigation strategies to ensure that changes to source data are quickly replicated across global sites. Skkyne's proprietary DataHub Transfer Protocol enables bidirectional low-latency data transfer; the provider worked with TEPSYS to digitize and centrally orchestrate data collection, storage and visualization workflows for a hydroelectric power plant in Japan.

## 4. Data storage management.

Manual data storage management is time-consuming and error-prone, and can result in the costly build-up of dark data. Data management solutions provide tools to configure or automate when to archive or delete data – and when to cache data for fast-access applications such as real-time AI/ML inferencing, decision-making analytics and dashboards. Cognite's Data Fusion is an end-to-end SaaS platform, optimizing storage costs and access latency without the need for customer intervention. Similarly, AVEVA offers a variety of data storage management tools across its product suite, including Data Hub and PI System.

## 5. Data cleaning, lineage and feature extraction.

Time-series data, event logs, images and written reports can contain errors from sensor drift or failure, or be incorrectly labelled during a manual upload process. Industrial data management solutions enable real-time data quality checks and enforce data chain-of-custody for streamlined auditing and troubleshooting. Additionally, they offer tools for extracting useful features from unstructured data and for improving the discoverability of images, documents and audio. Yokogawa RAP partnered with Radial SG to deploy the latter's Viewport software within Yokogawa's product portfolio. Viewport streamlines asset documentation review using AI-powered software to extract text, symbols and other features from unstructured data such as documents and P&IDs. Similarly, Hitachi Image Based Inspections offers electrical utilities the ability to automatically extract features of interest from vast quantities of images from fixed cameras, satellites and drone-based inspections, to detect asset defects and bring these to the attention of O&M teams.



Figure 5

## Capabilities Criteria For Industrial Data Management Solutions

Capability	Description
Direct industrial data acquisition	Ability to connect directly to IoT sensors, cameras, microphones, asset PLCs, distributed control systems (DCSs), MESs, or the simultaneous control and data acquisition (SCADA) system, via common industrial protocols such as OPC-UA/DA, MQTT or through representational state transfer application programming interfaces (REST APIs) or custom connectors.
Indirect industrial data acquisition	Ability to connect to existing data hubs, data aggregators and enterprise information technology systems. Rather than pulling data directly from sensors, cameras, microphones and PLCs, data management solutions with this capability will connect to historians, data lakes and EAM and ERP systems.
Single source of truth (SSOT)	Ability to create a catalogue of references to consistent, synchronized version data from a variety of enterprise sources. SSOTs do not modify the original data – transformations are performed by reference only.
Data storage management	Ability to configure rules for data storage prioritization, archive timelines and eventual deletion. Industrial data storage management processes may involve caching new data for use in real-time AI/ML inferencing or compressing old data and saving on slow-access, low-cost, off-site storage. Other activities include performing regular back-ups, and detecting and deleting duplicates and unwanted data. Such capabilities allow industrial firms to save on storage costs while maximizing data utility.
Data cleaning, lineage and feature extraction	Ability to improve data quality, track data origin and transformations and extract metrics and useful features from unstructured data. Lineage tracking provided by data management solutions help determine root causes of data quality issues, and automatic extraction of useful metrics and features from time-series data, images, video, audio and report documents helps improve data discoverability.
Data access management and security	Ability to robustly restrict data access to authorized personnel and to endure regular cyber attacks.
Data contextualization	Ability to create self-explaining, discoverable data packages. Data management solutions offering this functionality give engineers the ability to annotate P&IDs with notes and references to documents and images, tag time-series charts with temporal events and package a variety of data into a self-explaining, digestible payload for data scientists and enterprise decision-makers.
External data sharing and benchmarking	Ability to securely share operational technology (OT) data with industry peers for the purposes of comparison, benchmarking or as part of an industrial data marketplace. As data-driven industrial asset management tools such as asset performance management (APM), predictive maintenance (PdM) and asset health and failure monitoring systems become more valuable, so the need for asset-specific historical data grows. Such capabilities offer industrial firms performance benchmarking and additional revenue streams through data licensing.
Data analytics	Ability to provide advanced, off-the-shelf analytics for a broad range of use cases, covering predictive maintenance, anomaly detection, emissions reduction, production optimization and others.
Data visualization	Ability to visualize data lineage, data transformations and data analytics within the software platform. Such capabilities may include GUIs with interactive data manipulations, network diagrams, charts, dashboards and heatmaps.

Source: Verdantix analysis



## 6. Data access management and security.

Critical infrastructure, industrial facilities and enterprise systems today are subject to intense cyber attacks. The 2021 ransomware attack on Colonial Pipeline shut down the transportation of refined oil products for five days, resulting in a fuel shortage along the East Coast of the US (see [Verdantix Market Insight: The Convergence Of Physical And Digital Risk Management](#)). While industrial data management solutions that centralize access to disparate data sources can present additional cyber security concerns, the holistic view they provide can also facilitate more robust access management and intrusion monitoring. With better visibility into data chain-of-custody, data use cases and potential vulnerabilities, access to sensitive data can be securely limited to relevant stakeholders. C3 AI, Palantir and Skkyne offer robust cyber security features for their partnerships with government agencies, nuclear power plants and defence contractors.

## 7. Data contextualization.

Data management solutions such as DataOps platforms offer the ability to annotate P&IDs with notes and references to documents and images, tag time-series charts with temporal events and package a variety of data into a self-explaining, digestible payload for data scientists and enterprise decision-makers. Cognite worked with chemicals manufacturer Celanese to consolidate and contextualize equipment hierarchy, 3D data, unstructured documents and production data for the collaborative development of analytics and improved data discovery at a pilot facility in Clear Lake, Texas.

## 8. External data sharing and benchmarking.

Vendors with the ability to securely share OT data with industry peers for the purposes of comparison, benchmarking or as part of an industrial data marketplace help drive digitization across the industry as a whole. Only a handful of vendors currently offer this capability. AVEVA allows customers to securely share and benchmark OT data with peers as part of its Data Hub Community, while MachineMetrics analyses trends across its customers to provide predictive maintenance and production optimization.

## 9. Data analytics.

The whole purpose of establishing strong data management processes is to eventually use high-quality, contextualized industrial data to drive algorithms that transform this into actionable insights. The past decade has seen the rise of innumerable AI-based analytics providers offering out-of-the-box, quick-to-implement analytics for a wide variety of use cases, such as machine failure, supply chain disruption and production optimization. AspenTech scored the highest amongst all the vendors in the Verdantix Green Quadrant benchmark for asset failure prediction (see [Verdantix Green Quadrant: Asset Performance Management Solutions 2022](#)), while C3 AI offers analytics for a broad range of use cases, such as turnaround optimization, energy management, inventory optimization and ESG.

## 10. Data visualization.

Unless it is displayed in a human-readable format, the most advanced analytics in the world will do nothing to improve operational decision-making. From a simple signal vs time plot of time-series data, to a fully interactive three-dimensional heatmap projected onto realistic digital models of an industrial facility – data visualization is the critical cherry on top of a compelling industrial data management solutions portfolio. Speciality materials manufacturer Arkema implemented TrendMiner, a subsidiary of Software AG, to develop analytics and visualizations to provide insights specific to individual tasks.



Figure 6  
Vendor Capabilities: Industrial Data Management Solutions

	Direct Industrial Data Acquisition	Indirect Industrial Data Acquisition	Single Source Of Truth	Data Storage Management	Data Cleaning, Lineage And Feature Extraction	Data Access Management And Security	Contextualization	External Data Sharing And Benchmarking	Data Analytics <sup>1</sup>	Data Visualization
AspenTech	●	●	●	◐	◐	◐	◐	◐	◐	◐
AVEVA	●	●	◐	●	◐	◐	●	●	●	●
C3 AI	◐	●	◐	◐	●	●	◐	◐	●	●
Cognite	◐	●	◐	◐	●	◐	●	◐	◐	●
DataProphet	◐	◐	◐	◐	◐	◐	◐	○	◐	◐
Element Analytics	◐	●	◐	○	◐	◐	◐	○	○	◐
HighByte	●	◐	◐	○	◐	◐	◐	○	○	◐
Hitachi Vantara	●	●	●	●	●	◐	●	◐	◐	●
Honeywell	●	●	◐	◐	◐	◐	◐	◐	◐	◐
IBM	◐	◐	◐	◐	◐	◐	◐	◐	◐	◐
MachineMetrics	●	◐	◐	○	◐	◐	◐	◐	◐	◐
Palantir	○	◐	◐	○	◐	●	◐	○	◐	◐
PTC	◐	◐	◐	◐	◐	◐	◐	○	◐	◐
Radial SG	○	◐	○	○	◐	○	◐	○	◐	◐
Seeq	◐	◐	○	○	◐	◐	◐	◐	◐	●
Skkynt	◐	◐	◐	○	○	●	○	○	○	◐
SymphonyAI Industrial	◐	●	○	○	◐	◐	◐	○	◐	●
TrendMiner	◐	◐	○	○	◐	◐	◐	◐	◐	●
Uptake	●	●	●	◐	◐	◐	◐	○	◐	◐
YuzeData	◐	◐	◐	○	○	◐	◐	◐	◐	◐

No capability	○
Limited capabilities	◐
Some capabilities	◑
Strong capability	◒
Market-leading capability	●

Note: Capabilities based on publicly available case studies and vendor demonstrations, where applicable. <sup>1</sup> Refers to breadth of capabilities.  
Source: Verdantix analysis



# Customers Must Develop Vision And Strategy Before Beginning Their Data Management Journey

Holistic data management solutions are marketed extensively in the industrial technology world, with some vendors promising end-to-end solutions for their clients' every OT data management need. The reality is, as expected, more nuanced – and depends on the digital maturity of industrial operations and the variety of systems with which they can integrate. Industrial firms, regardless of size and digital maturity, must develop a clear vision of what they want to achieve through digitization, and how to get there. In addition, industrial firms should consider that:

- **Entrenched, legacy IT/OT infrastructure cannot be replaced overnight.**

Enterprises may rigorously plan ahead, consider the trajectories of contemporary technologies and attempt to ensure that future additions and upgrades to processes and structures run with minimal disruption. However, the unimaginably complex interrelated system of assets, workers and external conditions means that strategists will make mistakes. Add to this the pressure of maximizing operational efficiency – and the mentality of 'if it ain't broke, don't fix it' will begin to bury and ossify decades of legacy IT and OT infrastructure. Global industrial firms know this all too well, with 46% of our 2022 survey respondents seeing replacing old industrial equipment as a high-priority activity for 2023. Such equipment and systems may be integral to everyday operations, necessitating consultation with stakeholders across an enterprise to establish a strategy aligned with business resilience (see [Verdantix Best Practices: Transitioning To Predictive Maintenance For Enhanced Asset Management](#)).

- **Digital immaturity presents challenges with workforce integration.**

While a lack of legacy software may appear to benefit firms beginning their digitization journeys, a greater barrier may be presented by entrenched manual processes and an unwilling workforce. MachineMetrics worked with components manufacturer AccuRounds to provide connectivity to 23 CNC (computer numerical control) machines, with the aim of improving production efficiency in the context of a workforce uncomfortable with perceived increased scrutiny of their every move. To address worker friction, MachineMetrics installed monitors for machine operators to understand more about the monitoring process through intuitive, contextualized data visualization, and to help optimize their work tasks. As a result, worker turnover reduced by 52% and OEE increased by 20%.

- **Service providers bring vision and strategy to data-driven digitization.**

Consultants offer a holistic perspective and the experience necessary to bring together relevant functions, disparate technologies and systems into a single-pane-of-glass view for decision-makers. Half of the 301 respondents in the 2022 Verdantix global corporate survey mentioned plans to use third-party consultants for industrial data management and governance in 2023 (see [Verdantix Market Insight: Industrial Asset Management Digital Services – Spending Plans And Use Cases By Industry](#)). In October 2022 Accenture began working on a cloud platform for data and AI with food and beverage manufacturer Mars, to enable real-time insights for line operators, increase the scope of automation and improve sustainability.

## Vendors Should Expand Data Acquisition Towards Varied, External Sources

Industrial firms will develop strategies to replace legacy data management hardware and software systems through workforce consultation and integration. Particularly challenging digitization projects will see input from asset management service providers, who may choose a composite solution through the integration of multiple software vendors. To maximize return on investment (ROI) for customers, industrial data management solution providers should enhance their offerings to encompass:



- **AI-powered knowledge engines to clarify the waters of firms' data lakes.**

AI and ML techniques such as NLP have made giant leaps in recent years. In 2019 Google began using its BERT model to improve natural language understanding for Search queries, while in January 2023 Microsoft announced a multi-billion-dollar investment in OpenAI as part of a focus on consumer and enterprise-oriented generative AI and more powerful search engines. Data management solution vendors should explore NLP-powered tools for improving the ease of use of data lake queries and enhancing the discoverability of text-based content using semantic similarity. In addition to NLP, vendors should build out functionality for extracting semantic information from P&IDs, images and video through ML-based OCR and image classification. Such AI-powered knowledge engines will allow data scientists and decision-makers to find relevant data more quickly, with fewer queries – while less code-heavy interfaces will provide machine operators and maintenance personnel with a more accessible portal into critical industrial data. NLP-focused search developer deepset worked with Etalab to help replace keyword-search systems for French government databases with deepset's Haystack – an open source, AI-focused question-answering system for enterprise-scale document data sets.

- **Emissions data for compliance and reduction.**

In March 2022 the US Securities and Exchange Commission (SEC) published its Climate Disclosure Rule, which may come into effect in FY 2023 (see [Verdantix Strategic Focus: Complying With The SEC Climate Disclosure Rule](#)). While specifics are still evolving, industrial firms must begin to establish emissions data management across their supply chains – defined by Scope 1, 2 and 3. Carbon management software providers have such capabilities (see [Verdantix Green Quadrant: Enterprise Carbon Management Software 2022](#)), but DataOps platform providers may offer additional OT-focused, application-specific features, such as data cleaning and contextualization – helping to fast-track emissions reduction strategies. The next 12-24 months will witness an increase in collaboration between industrial DataOps platform providers and carbon management software vendors or emissions reduction analytics specialists.

- **Benchmarking capabilities against industry peers to democratize best practices.**

As old, inefficient industrial equipment is replaced or retrofitted with new parts, sensors and control systems, firms will lack relevant historical data. Software vendors offering the ability for customers to benchmark the performance of their new assets against peers with similar systems will improve competition and help the industry as a whole reduce downtime and boost production efficiency. AVEVA worked with automotive parts manufacturer HENN to discover and contextualize performance data from a range of suppliers to provide clear comparisons across several metrics. Each supplier was able to see the performance of its peers, which increased competition and resulted in HENN achieving a 10% improvement in overall equipment efficiency.

- **Advanced visualizations.**

Distilling complex operational data to maximize the clarity of features relevant to decision-making is critical to industrial DataOps. For example, a spectrogram converts an analogue vibration into a temporal heatmap, allowing fine details recorded from microphones to be displayed in an uncluttered, human-readable format. Additional visualizations include configurable annotations over charts, where spectrograms can be labelled to indicate unusual machine noises and noteworthy findings from asset inspections. More advanced visualizations, such as richly visual 3D digital twins of industrial facilities, will allow O&M teams to quickly navigate sites and review issues according to colour-coded priorities (see [Verdantix Applications Of 3D Visualization Software And Its Benefits For Industrial Asset Management And Operations](#)).



# Independent insight and analysis

Our research is a trusted source for some of the largest and most innovative businesses in the world. With over a decade of reports, data and analysis, our subscribers have access to depths of insight that cannot be found elsewhere.

Whether you are implementing a leading-edge technology strategy, or developing the products and value propositions of the future, our analysis will help you futureproof your thinking.

## Our expertise

Environment, Health & Safety

ESG & Sustainability

Net Zero & Climate Risk

Operational Excellence

Smart Buildings

## Contact

Verdantix Ltd, 18 Hatfields, London  
SE18DJ, United Kingdom

[contact@verdantix.com](mailto:contact@verdantix.com)  
[@Verdantix](https://www.verdantix.com)

## Opportunities at Verdantix

Since 2009, Verdantix has been delivering high-quality research and advice to its clients. If you're interested in joining a world-class team with an unwavering focus on success, apply to join us today. We are delighted to be hiring across all teams and have a variety of opportunities in both London and Boston

