COGNITE

Data Contextualization for Industry

Convert data to knowledge, deploy and maintain solutions at scale

Data Contextualization for Industry

Convert data to knowledge, deploy and maintain solutions at scale

"Data has no value unless the business trusts it and uses it".

FORRESTER®

About Cognite

Cognite is a globally recognized leader in industrial software with a clear vision: to rapidly empower industrial companies with accessible, trustworthy, and contextualized data and drive the full-scale digital transformation of asset-heavy industries.

With its market-leading Industrial DataOps platform, **Cognite Data Fusion®**, and a comprehensive suite of Industrial generative AI capabilities, Cognite AI, Cognite makes it easy for decision-makers to access and understand complex industrial data. **Cognite Data Fusion®** is a user-friendly, secure, and scalable platform that enables industrial data and domain users to collaborate quickly and safely to develop, deploy, and scale industrial generative AI solutions that deliver both profitability and sustainability.

WHY COGNITE DATA FUSION® →

BENEFITS FOR YOUR TEAMS →

CONTACT SALES →

Table of contents

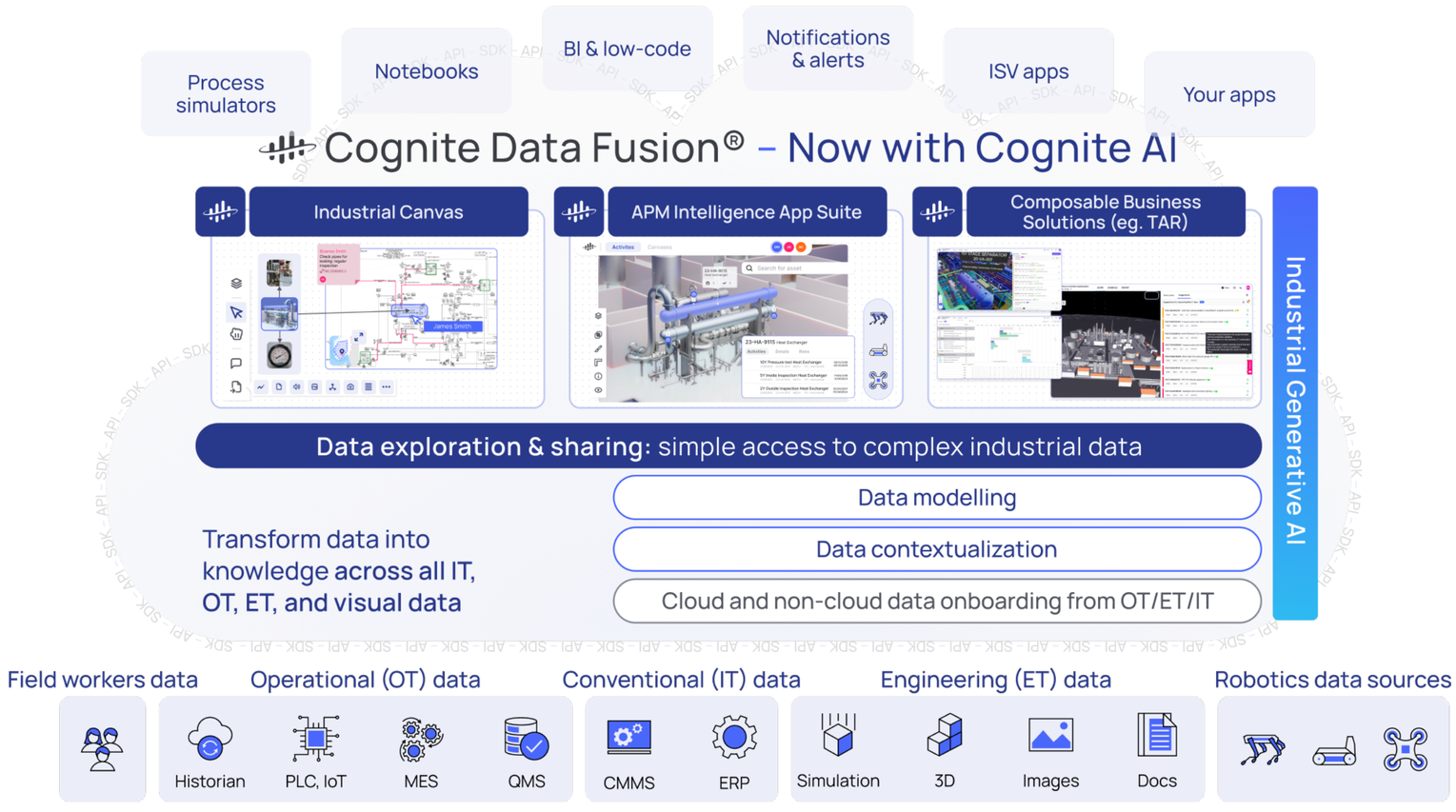
Simple access to complex industrial data

For every one person who can 'speak code,' there are 100s of others who do not, especially in the industrial environments where there are numerous data types and source system complexity. **To use industrial data broadly, it requires context.**

Subject matter experts, field engineers, and data scientists deserve simple access to all industrial data in a single workspace. This requires a unique way to leverage and apply contextualized data. Industrial applications, as we know them today—especially data dashboarding and visualization—will soon be 100% transformed. Open, composable workspaces with integrated AI Copilots will become the point-of-entry to engage with industrial data in the same way the web browser replaced desktop applications, while the infusion of generative AI accelerates the ability to converse with contextualized, trusted data via a performant API – without writing a single line of code.

Cognite is on a mission to empower industrial organizations with **simple access to complex industrial data.** The Industrial DataOps foundation provided by Cognite Data Fusion®, coupled with the power of generative AI and data modeling, derives hidden relationships for cross-data source insights. It's about taking away the coding and scripting and instead, about creating an intuitive end-user experience to generate operational improvements.

▼ Figure 1: Cognite Data Fusion's simplified service architecture/offering



WHY COGNITE DATA FUSION® →

BENEFITS FOR YOUR TEAMS →

CONTACT SALES →

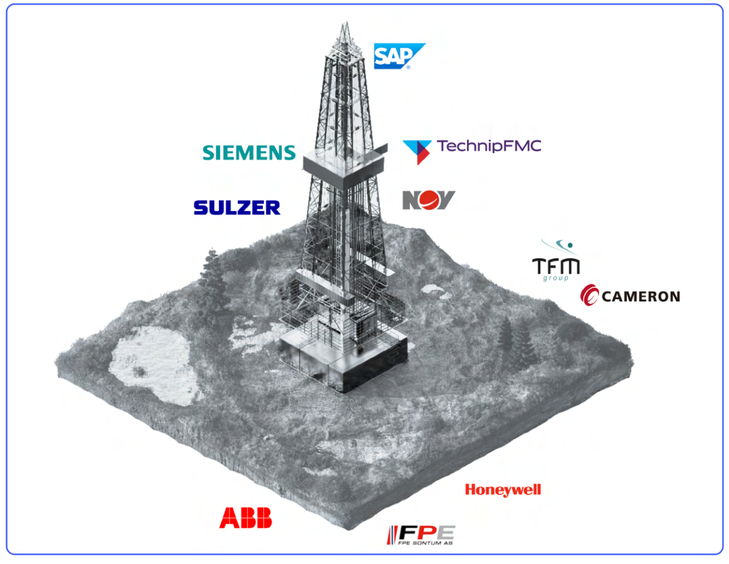
The Industrial Data and AI problem

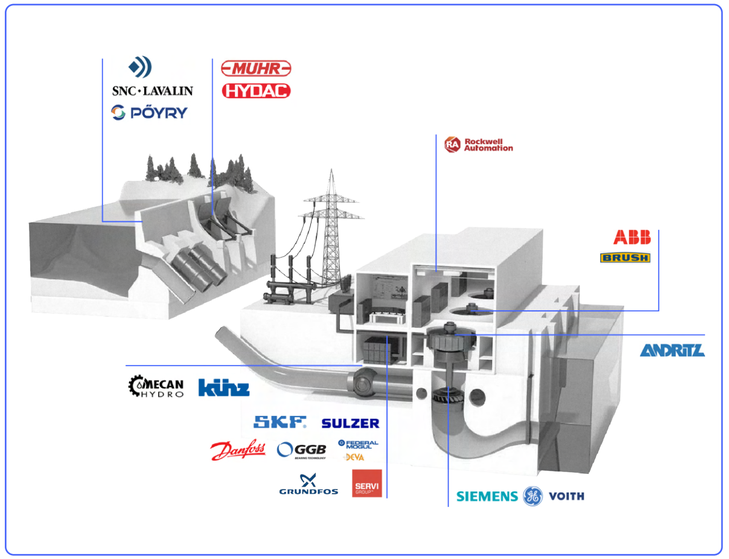
Drowning in data, starving for context

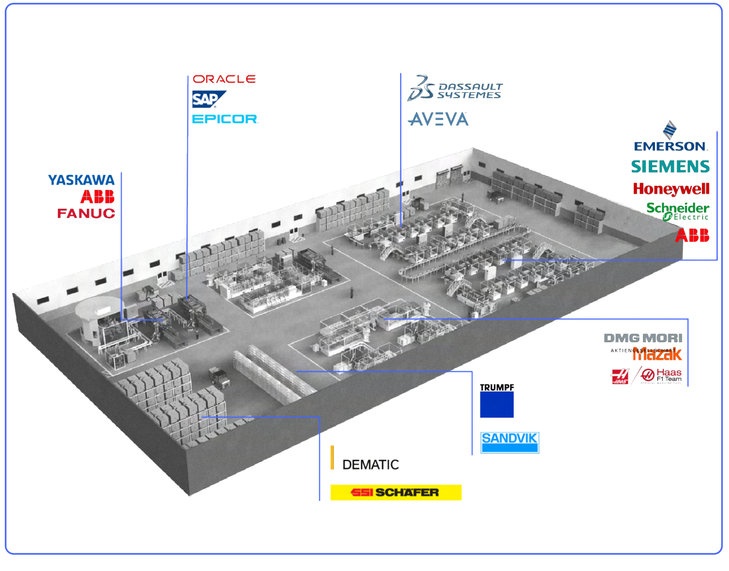
For a typical industrial facility, more than 100,000 data points are routinely generated across more than 50 applications (Fig. 2). Unstructured data types are on the rise with more reliance on images, videos, acoustic, 3D models, pointclouds, and engineering drawings to provide additional context on the state of operations. Traditional efforts to connect data from systems are manual, time consuming, and fail to manage structured data at scale, much less incorporating the growing unstructured data.

Enterprises large and small are rushing to reduce the barriers their workforces must overcome to consume data — or to be more data literate. Gartner formally defines data literacy as "the ability to read, write and communicate data in context," more informally expressed as "Do you speak data?". Data literacy includes an understanding of data sources and constructs, analytical methods and techniques applied to data, and the ability to describe the use case applications and resulting value.

The need to understand industrial data is also relevant to the growing demand to apply generative AI in these environments, where large language models (LLMs) like ChatGPT lack the industrial context required to provide deterministic, trusted responses unique to each facility.







▶ Figure 2: Examples of typical industrial facilities with data generated from various systems

In the past, people were forced to “speak data” in order to gain actionable industrial insight. Now, we have progressed to make industrial data speak human. Complex industrial data can be liberated and put into context for the people who need it.

Going forward, we must ensure that your AI investments speak to your people about your industrial data. Those who can already provide simple access to complex industrial data are already delivering more business value with an ROI above 400% → and are well-positioned to rapidly make use of large language models to unlock new business opportunities.

While the urgency for subject matter experts (SMEs) to become more data literate is clear, they are faced with the industrial data and AI problem. The physical world of industrial data is a messy place. As an example, equipment wear, fluctuating operational targets, and work orders are all important when assessing the root cause of an operational issue. A process variable cannot be observed and interpreted in isolation. Is the variable trending as expected? What is the maximum value recommended by the vendor? When was this equipment last inspected and what were the observations?

These questions concern SMEs and field engineers responsible for keeping equipment running and continuously operating at a level that optimizes short-term and long-term production efficiently. Increasingly, documentation and electronic trails of maintenance and operational history can be accessed digitally. However, diagrams and vendor documentation are still found in flat PDF files, maintenance records are scanned paper documents, and 3D models are not up-to-date with the actual physical world as modifications and maintenance have occurred over the asset’s lifetime. How do we enable SMEs across operations, maintenance, and reliability to solve their problems with digital tools?

To adequately empower SMEs to extract the value of industrial data insights, operationalizing data must become a core part of your business strategy. Data must be liberated, contextualized, and easy to use to generate data-driven insights. A robust data foundation powered by a solid contextualization engine is vital for enhancing operational efficiency. This provides a solid base for running generative AI models, helping accelerate workflows, automate repetitive tasks, assist in decision-making, and more.

It’s all about relationships

> “Contextualization of data means to uncover and identify relationships between elements of data that are connected, but where the relationships are not explicitly represented. This can be done in many ways depending on the type of data and the kind of relationship - some of the techniques are simple pattern matching, others can rely on being able to understand specific data formats, having a lot of domain knowledge, an ability to identify patterns that are not exact matches.”

The value of data contextualization is the automated discovery of relationships in diverse data, transforming data into knowledge for a rapidly growing data consumer and generative AI landscape. Indeed, the ability to identify meaningful relationships — across data types, people, places, and objects — is absolutely fundamental to faster generation of real value from data and analytics.

Before

Do you "speak data"?

Now

Does your data speak human?

Next

Does your AI speak your data?

WHY COGNITE DATA FUSION® →

BENEFITS FOR YOUR TEAMS →

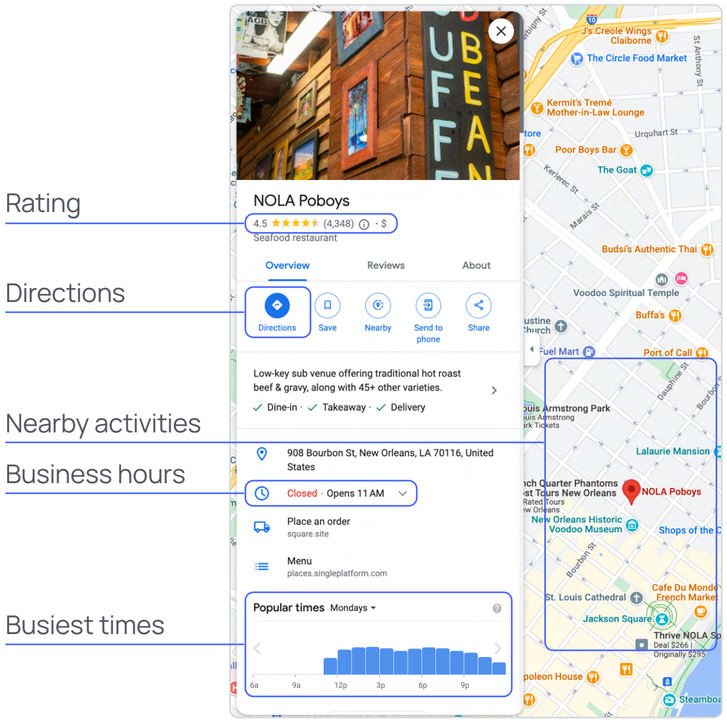
CONTACT SALES →

Think, for example, on what makes Google Maps effective. Google Maps combines map data with information from the web (rating, direction, hours of operation, nearby activities, busiest times) to provide users with a unified interface to quickly answer multiple questions from a single, seamless experience (Fig. 3). In many cases, these queries don't even need to be complete to provide useful information and quick answers to questions. Instead, the user can navigate the data in context to take an informed action.

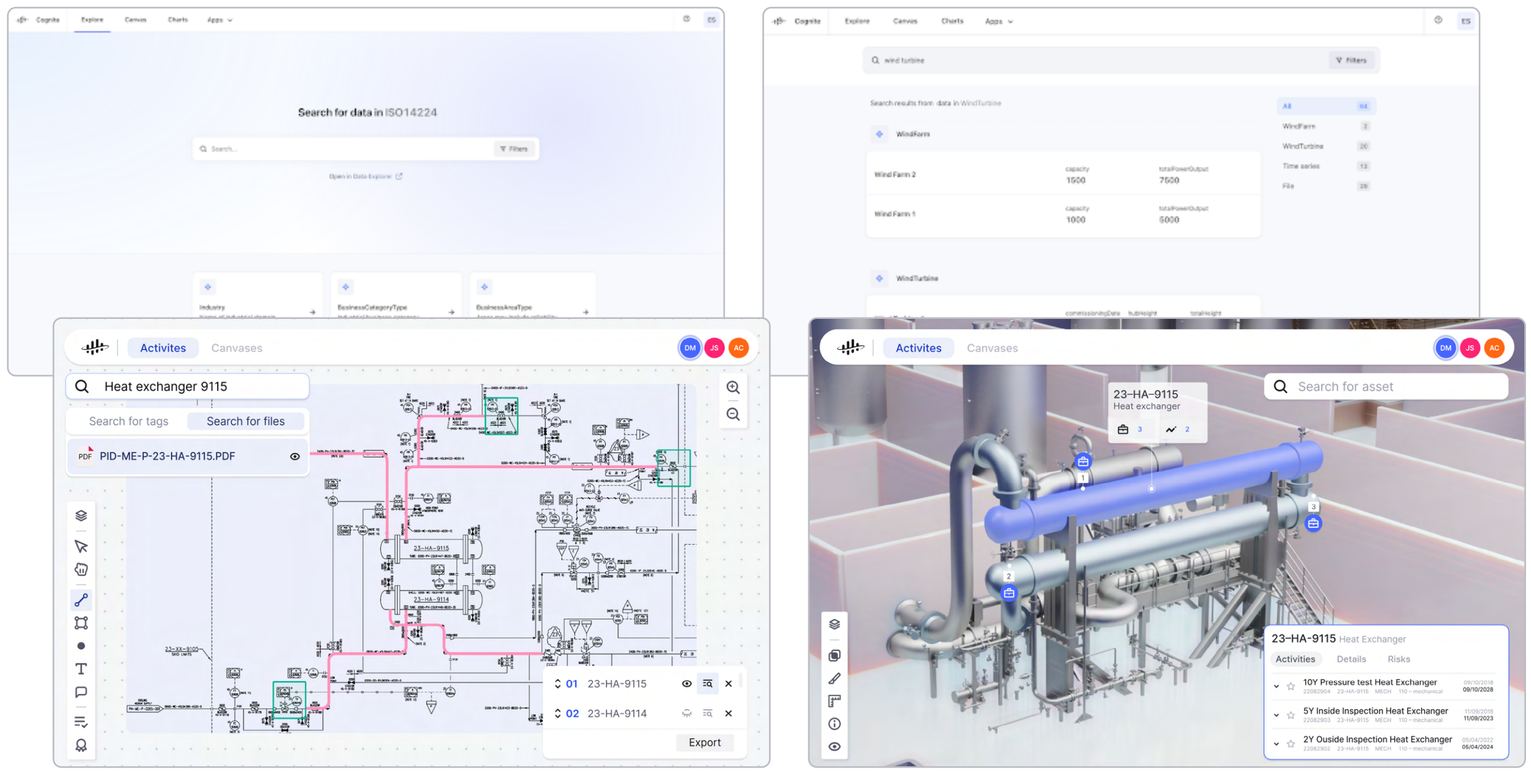
But what about upstream energy production, downstream and process manufacturing, hybrid and discrete manufacturing, power generation, and other industrial environments?

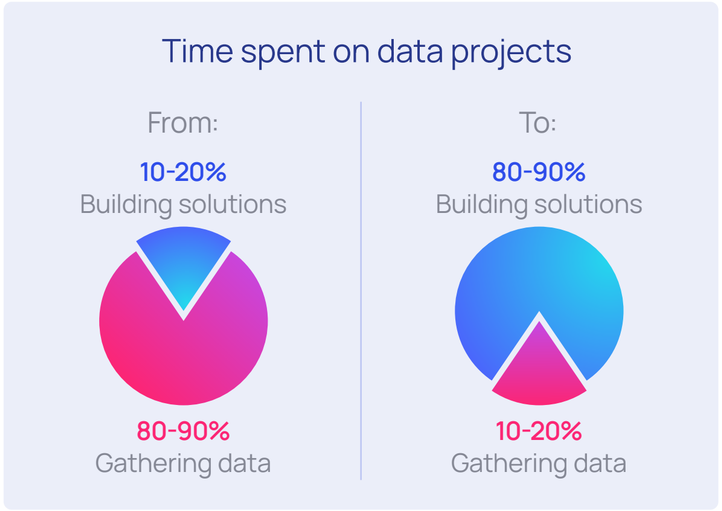
Given the high-stakes nature of these industries and the sheer amount of operational questions being asked, operators in these environments deserve to easily navigate all related data in a Google-like search, enhanced by generative AI, to make faster decisions, ensure safer working conditions, and improve asset reliability and resiliency of operations.

Although consumer technologies and industrial technologies are different, the approach is quite similar. Similar to Google Maps, Cognite Data Fusion® puts structured and unstructured data, real-time and historical, into business context, enabling anyone to build, deploy, and scale digital solutions that drive business value. You can start navigating contextualized information with a Google-like search, a 3D view, or a drawing (P&IDs or process flow diagrams) in order to find the data you need (Fig. 4).



▲ Figure 4: Example of navigating equipment's contextualized information in Cognite Data Fusion®





▲ Figure 5

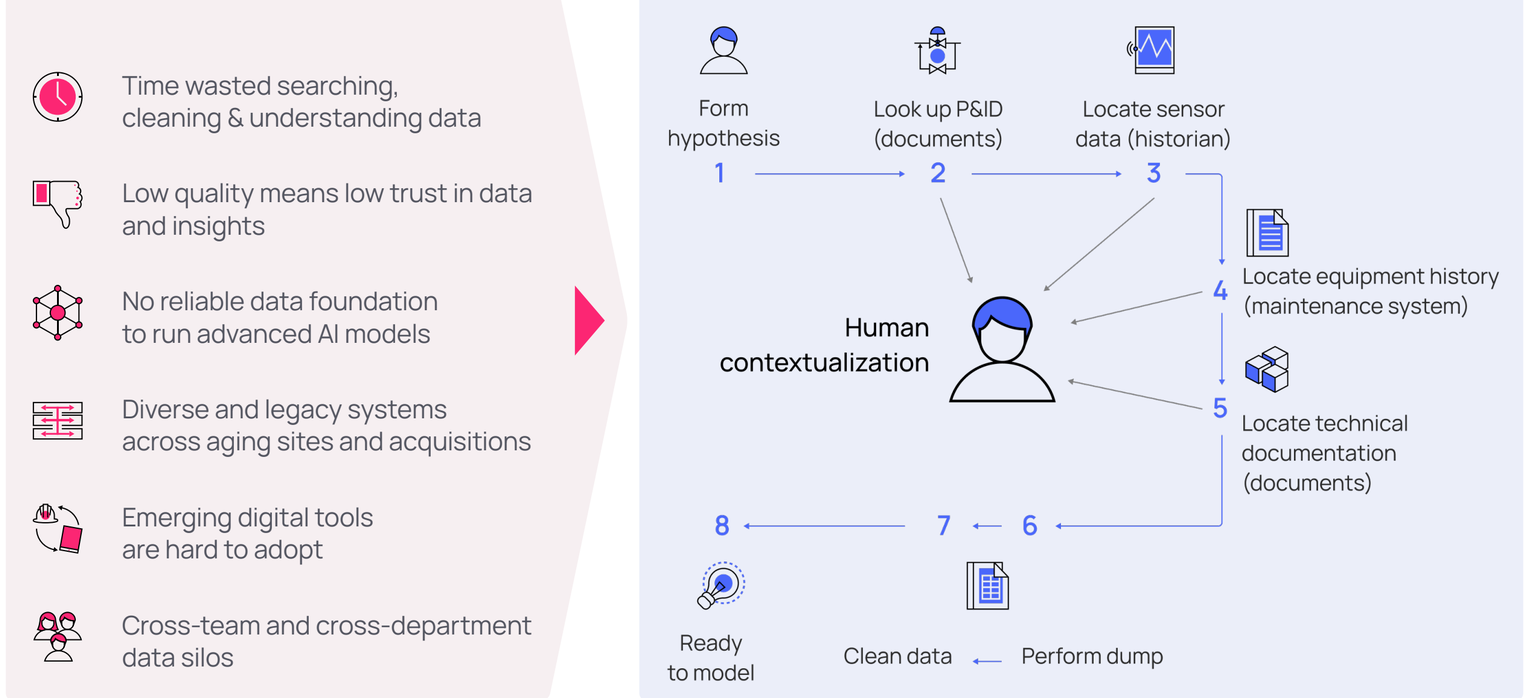
No matter the point of entry to being navigating data,, the result is the same: the day-to-day life of the user becomes much easier. Instead of subject matter experts spending 80% of time finding and aggregating information, contextualized data turns that around and empowers end users to spend less time gathering data and more time focused on making quicker, better-informed decisions, with the help of generative AI and copilots (Fig. 5).

Unfortunately, today's reality for industrial search is too complex. SMEs spend hours looking for and interpreting data they need to tell an asset-centric narrative, just to learn that the data is incomplete, inaccurate, or that the decision no longer matters. This can involve looking through reports and spreadsheets, engaging with other data owners, or making new requests to IT – none of it is happening through a single pane of glass (Fig. 6).

While there have been many previous attempts to solve the industrial data and AI problem with solutions like data lakes, these solutions lack the relationships between all data in the lake. Data lakes are not user friendly enough for SMEs, field engineers, and operators to quickly find the data they are looking for. They also lack the necessary context for generative AI solutions to provide deterministic answers to questions, resulting in a high probability of hallucinations and low levels of trust.

To solve the industrial data and AI problem, data must be put in industrial context. Instead of using SME brain power and sheer will to build the relationships that matter between these different data sources, industry must turn to algorithms and matching technology that not only springboard the mental model, but learn and can automate data relationship matching over time, enabling AI-powered business decisions.

▼ Figure 6: Example of a manual and siloed workflow caused by inability access industrial data

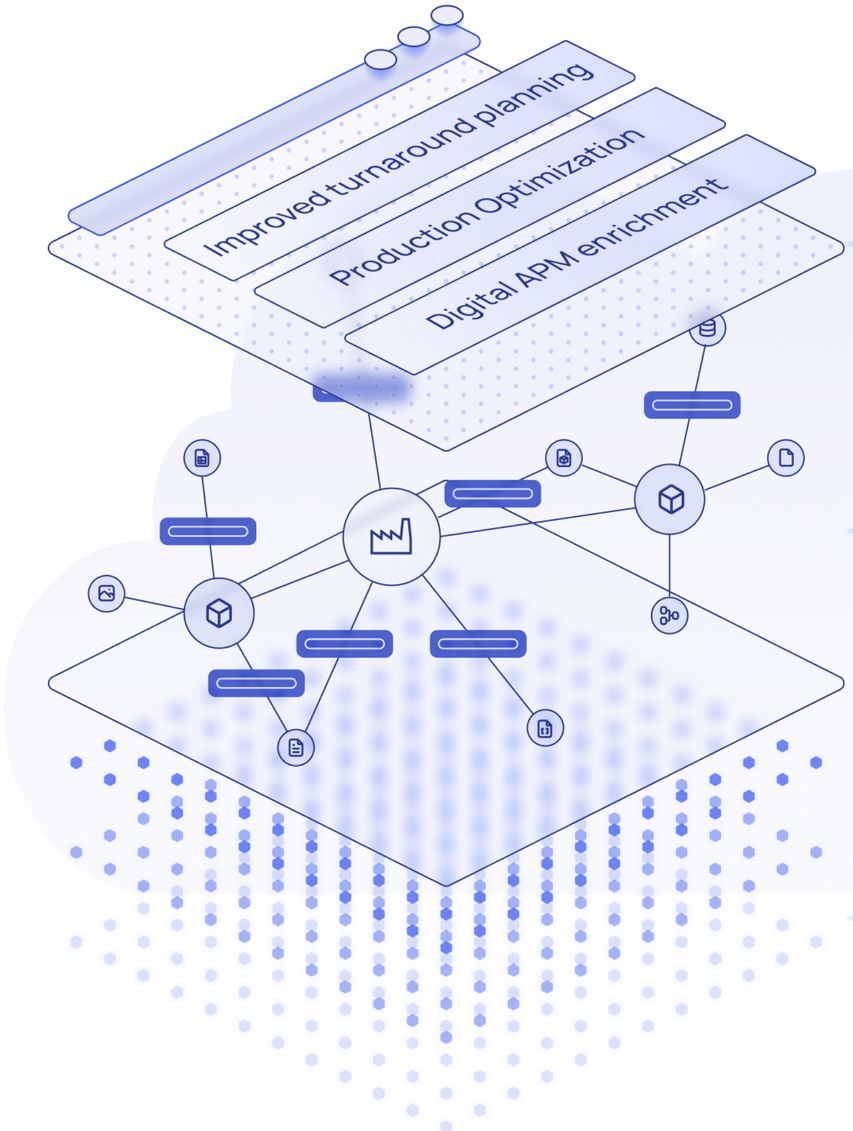


Data Contextualization as a foundation for innovation

Contextualization is the process of establishing meaningful relationships between data sources and types to traverse and find data through a digital representation of the assets and processes that exist in the physical world. Continually contextualizing disparate data sources is an iterative process that creates a rich data foundation for operational innovation.

As relationships between previously siloed data sources are established in this data foundation, you naturally start building an **industrial knowledge graph** tailored to your operations. The knowledge graph is continuously evolves and spans many dimensions and different data types, from the time series values to diagrams showing the process flows to a point cloud 3D model with recent images from an inspection.

Data from the industrial knowledge graph is deterministic, trustworthy, and accessible through a robust and performant API. These characteristics make the industrial knowledge graph necessary for **enabling generative AI use cases, modeling data for open industrial digital twin use cases, and providing an open workspace to generate insights with contextualized data (for example, in an Industrial Canvas)** (Fig. 7). Each of these topics will be discussed in detail, but first, we will start with how data is contextualized within Cognite Data Fusion®.



Industrial Canvas

Your teams deserve the industrial data & analytics UX only available with Industrial Canvas

- Simple access to complex industrial data in one collaborative workspace
- Use AI to develop and analyze complex cross data-source scenarios 90% faster
- Trouble shoot with real-time data and collaborate on RCA across teams and even sites

Open Industrial Digital Twins & Industrial knowledge graph

Represent our full enterprise as data graph with Open Industrial Digital Twins

- Build dynamic models of physical assets and processes, delivering real-time intelligence for active decision making
- Use data product practices to provide simple access with strong governance across source, domain, and solution data models across your ecosystem
- Take advantage of LLMs with Retrieval-Augmented Generation (RAG) powered by your own proprietary data, without data leakage

Scalable Industrial DataOps foundation

Supercharge your data engineering with Industrial DataOps

- 100+ OT, ET, IT, Simulator, and Robotics connectors out-of-the-box
- Connect from edge to cloud with confidence, with granular access controls, data lineage, and quality assurance built-in
- Catalog and activate you data across source, domain, and solution data models across your enterprise

▲ Figure 7: Capabilities enabled by Cognite's Data Fusion's industrial knowledge graph

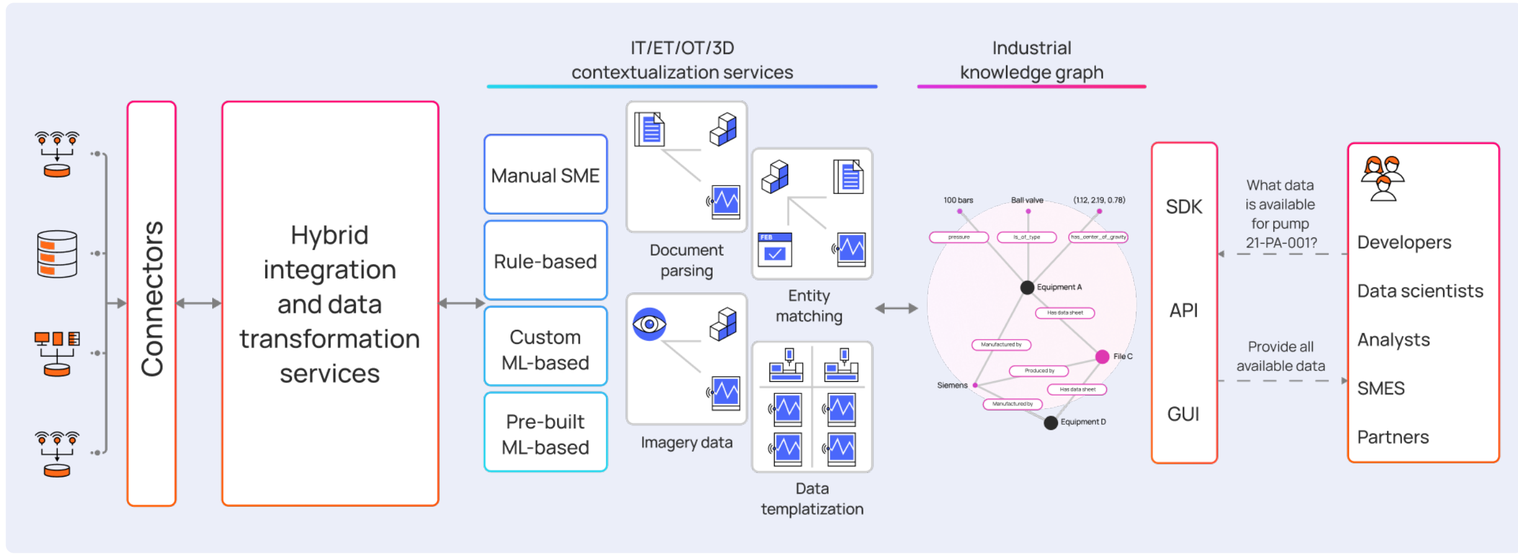
Cognite Data Fusion's Contextualization Engine

Building contextualization pipelines

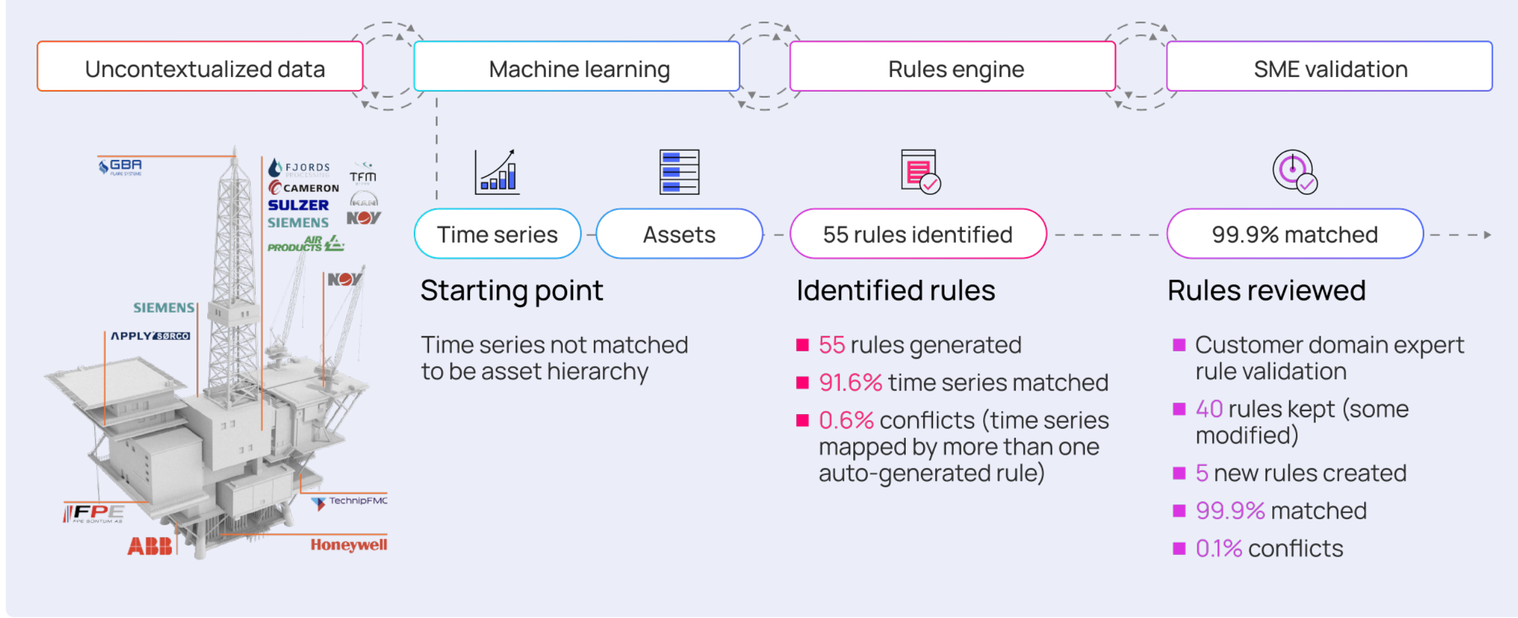
Data contextualization involves connecting all the data to better understand an asset or facility. In Cognite Data Fusion® , data relationships are created through contextualization pipelines that develop and maintain a comprehensive, dynamic industrial knowledge graph. This process involves connecting all OT, IT, and engineering data types and reusing the industrial knowledge graph across many business solutions and business domains (Fig. 8).

Cognite Data Fusion® delivers contextualized data as a service via a combination of AI-powered contextualization services using **pre-trained ML-based models, custom ML-based models, a rules engine, and manual/expert-sourced mappings**.

AI-driven contextualization permeates industrial data management, shifting the emphasis from data storage and cataloging to a true human data discovery experience, assisted further by a generative AI Copilot (Fig. 9).



▲ Figure 8: Data pipeline in Cognite Data Fusion



▲ Figure 9: Cognite Data Fusion's contextualization services/workflow

Cognite Data Fusion's contextualization capabilities extend to the following types of contextualization:

- Entity matching - map and connect time series, events, and tabular data to assets
- Create interactive diagrams - map and connect tags on engineering diagrams (P&IDs or PFDs) to other data sources
- Contextualize imagery data - upload and extract information from imagery data
- Create document classifieds - label training data and create document classifiers.

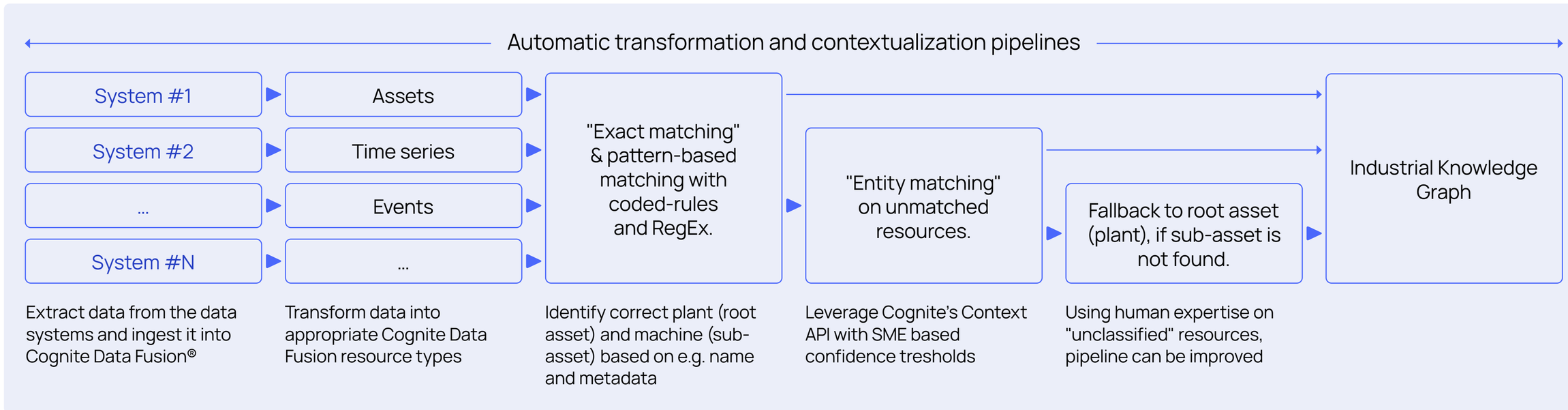
Entity matching

Cognite Data Fusion's contextualization capabilities match entities originating from various source systems. For different data sets to be successfully matched and appended into the reference and application data models, matching signals need to be present. The entity matching model uses AI to find matches when there are similarities between the strings and does not return suggested matches for unrelated entities. Even with weaker matching signals, a data contextualization engine provides enormous value by organizing, structuring, and governing an SME's intensive data contextualization work. For example, spreadsheets and CSV files are no longer needed to map relationships manually (Fig. 10).

Create interactive diagrams

The contextualization capabilities extend beyond entities and can also build interactive engineering diagrams/P&IDs from static PDF source files. Cognite Data Fusion® finds relationships between resources to set up, automate and validate all your contextualization pipelines from your browser without writing any code. With the mapped information, components are available to build applications where users, for example, can search a 3D model with the AI Copilot to quickly find an asset and see all the related time series data, or ask for all the pressure readings along a flow line.

▼ Figure 10: Example of how "entity matching" is leveraged in a contextualization pipeline to automate the formation of the Industrial Knowledge Graph



Contextualize imagery data

Whether captured with cameras, drones, or a mobile phone, images and video data contain valuable information about the state of a facility over time (inspections, maintenance routines, etc.). Yet, often utilizing this data remains a challenge. Cognite Data Fusion® also consists of several industry relevant and ready-to-use services for extracting information imagery data. For example, computer vision can identify objects like people, safety equipment spills, reading analog gauges, and much more. These services exist both as API and SDKs and can be used in automated pipelines for analyzing imagery data with the power of AI.

Create Document Classifiers

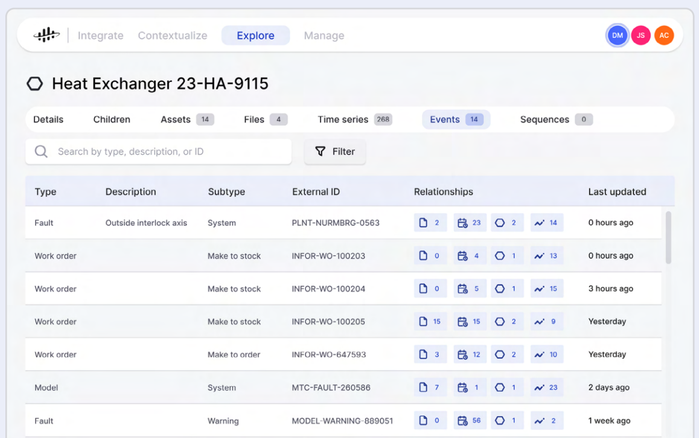
We can enrich documents such as standard operating procedures, inspection rounds, or OEM manuals with more context through connections to data sources like Sharepoint and local file directories. This contextualization will transform static pages into dynamic, interactive documents with live links to assets or processes identified in a document. Plus, contextualization enhanced by generative AI enables natural language document search, enabling users to quickly find the right information without having to manually scan through documents.

Cognite Data Fusion's unique ability to automate the contextualization process with AI-based algorithms across all of these different data types shortens the data contextualization process from months to days (Fig. 11). Using AI-based algorithms to efficiently build an industrial knowledge graph avoids 6+ months (based on Cognite's work with clients) of upfront efforts to give data context before any business value is achieved. Additionally, because these relationships are maintained as live pipelines, this effort can be managed at scale and data can be applied to solve many use cases, without having to dramatically increase the size of your team to manage this foundation.

▼ Figure 11: Examples of contextualized data sources

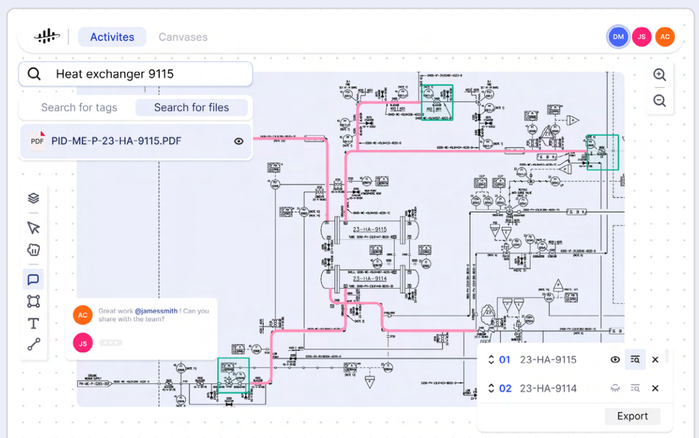
Contextualize events to assets

e.g. connect shop orders, work orders, and alarms to correct site, production line, and equipment



Contextualized documentation

e.g. connect tags in P&IDs to associated assets, time series, work orders, and documents etc.



Contextualized 3D models

e.g. filter for and visualize where work orders are located on plant

WHY COGNITE DATA FUSION® →

BENEFITS FOR YOUR TEAMS →

CONTACT SALES →

Contextualizing data into an Industrial Knowledge Graph

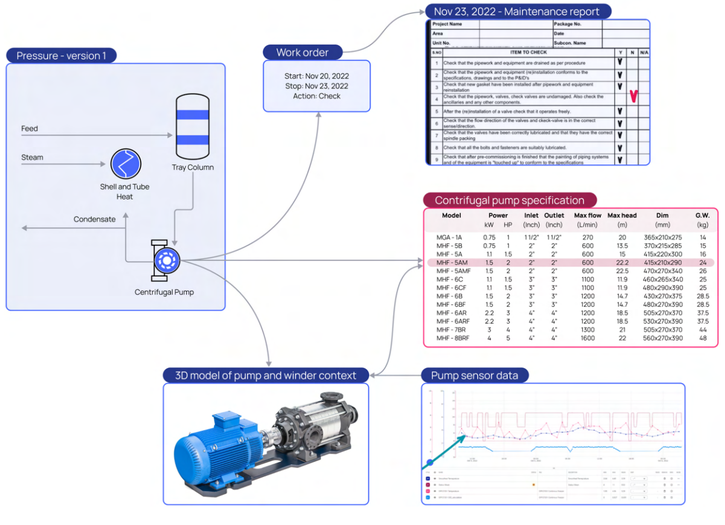
Cognite Data Fusion® is based on an open, flexible labeled industrial knowledge graph to represent your operations. Just like the principle of compounding interest, data in the industrial knowledge graph becomes increasingly more reliable the more people use, leverage, and enrich that data. More valuable and high-quality data leads to more trusted insights. More trusted insights lead to higher levels of adoption by your teams across the enterprise. A user-friendly, AI-powered experience also ensures user growth and adoption, and this cycle repeats exponentially.

Data modeling makes it easier for developers, data architects, business analysts, and other stakeholders to view and understand relationships between data objects. Take, for example, someone developing a production optimization application across multiple production lines. They need is a robust domain API that provides them instant access to a data model, which contains all the relevant data structured in a way that represents the process with performant querying, regardless of where that data originates from or where it is stored now. And they need it all in domain language, not in the language of databases. So in our particular, simplified example, what powers this domain API is the data model.

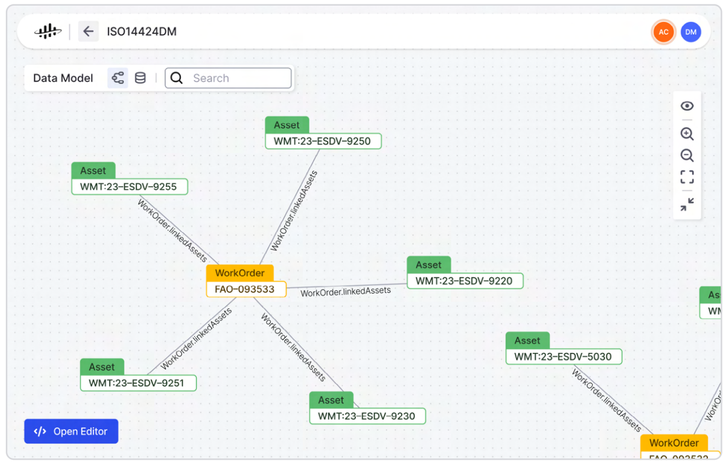
Figure 12 illustrates a simplified version of an industrial knowledge graph of a centrifugal pump.

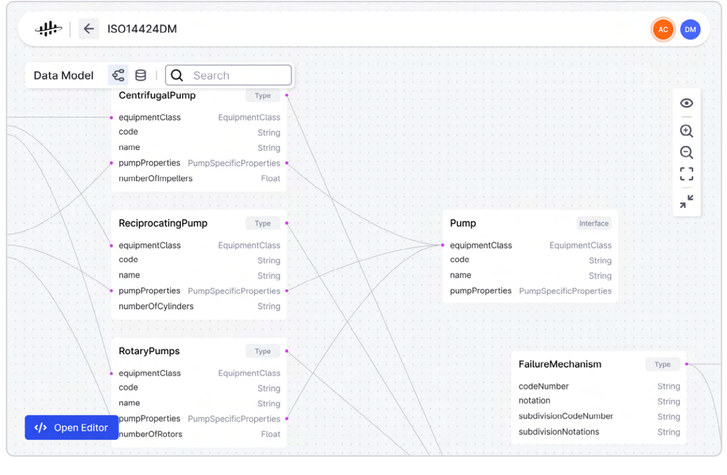
Depending on the persona, users may explore a problem with the pump from multiple entry points. Maintenance may start with the latest maintenance report, while an operator may use the time series, and a remote SME may begin with the engineering diagram (e.g., P&ID). The maintenance report, the work order, and the time series sensor values are each in separate systems; it is not a trivial task to gather all the relevant information necessary. This simple example illustrates the importance of data contextualization across different systems. Infused with generative AI, Cognite's data contextualization capabilities power the industrial knowledge graph (as seen on Fig. 13) to access the maintenance report, work order, time series, and more in a single location.

▼ Figure 12: Simplified industrial knowledge graph of a centrifugal pump



▼ Figure 13: AI-powered Cognite Data Fusion's Industrial Knnowledge Graph





WHY COGNITE DATA FUSION® →

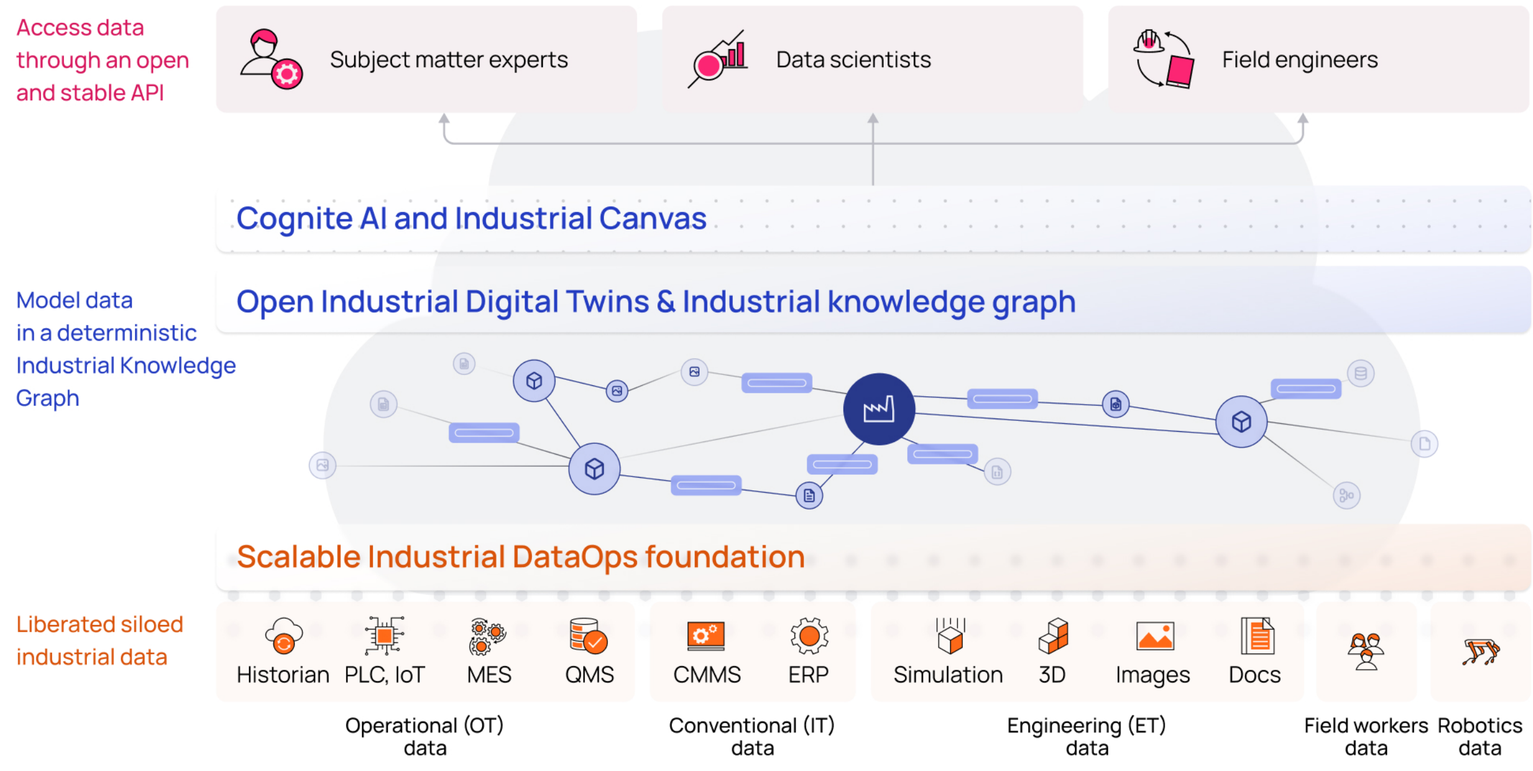
BENEFITS FOR YOUR TEAMS →

CONTACT SALES →

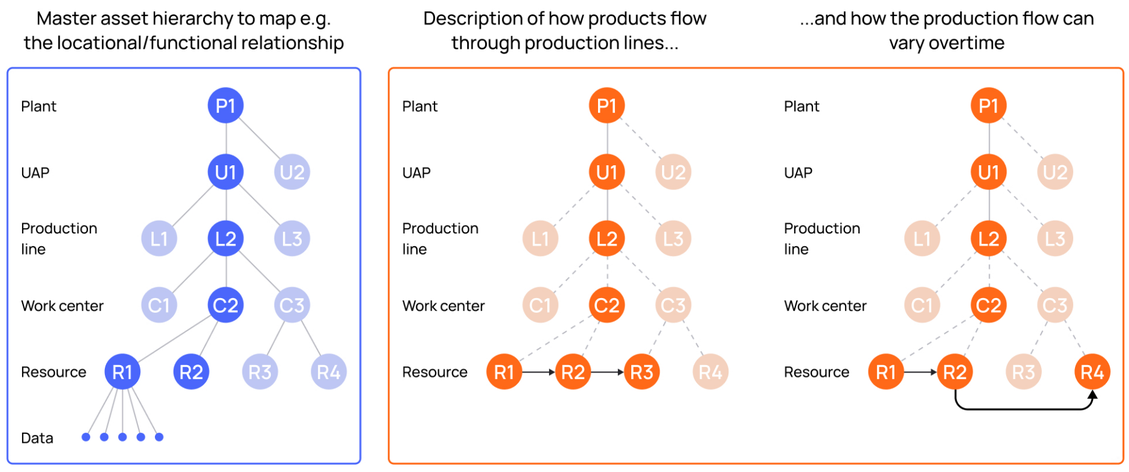
Simple aggregation of digitized industrial data is a significant step forward from the silos and inaccessibility that often plague large enterprises. However, to provide simple access to complex data, we need to account for the variety of industrial data types and incorporate the semantic relationships that drive scalable utilization of this data in support of interactive user experiences. Codifying this context in the form of an industrial knowledge graph is vital to enabling consistent, deterministic navigation of these meaningful relationships (Fig. 14).

The industrial knowledge graph provides the foundation for user-friendly data modeling. In the industrial knowledge graph, Cognite Data Fusion® stores IT, OT, and engineering data as Cognite resource types and builds meaningful relationships between the data so it is easy to understand what each data point is and how it relates to other data.

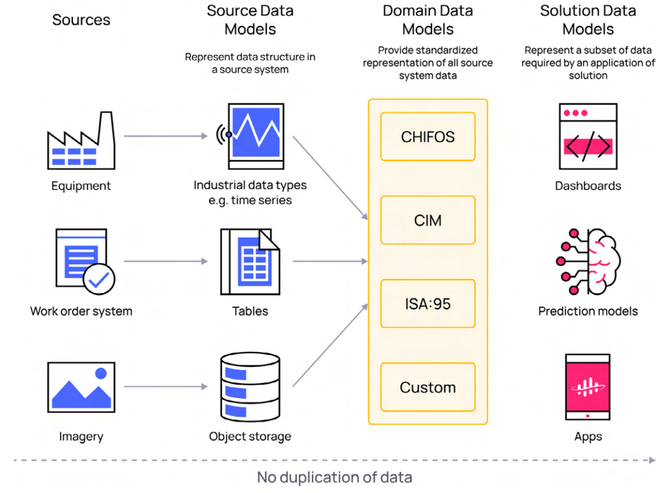
A data model in Cognite Data Fusion® is contextualized by establishing associations between related resources. These relationships are identified as resource types used to model associations between any two other resource types. It is through the notion of Relationships that graph-based data models can be represented in Cognite Data Fusion®.



▶ Figure 14: This is how Cognite Data Fusion's provides simple access to complex industrial data to subject matter experts, and continuously enriching this data via an industrial knowledge graph, which is available to everyone across the domains.



▲ Figure 15: The knowledge graph enables you to capture various data perspectives



▲ Figure 16: Three levels of data modeling

Typically, asset resources originate from a maintenance system, and the hierarchical structure from the maintenance system can define how the asset resources are organized in Cognite Data Fusion®. The relationships resource type allows you to organize the assets in other structures besides the standard hierarchical asset structure. For example, you can organize the assets by their physical location, where the grouping nodes in the hierarchy are buildings and floors rather than systems and functions. Another example is building a model that allows you to navigate assets by mimicking their physical connections through the production process (Fig. 15).

Data becomes an asset, liberated from its silos, with reusable analytics and scalable models, shareable across many users. This industrial knowledge graph encourages data reuse by creating a user-friendly architecture for your team. By leveraging data effectively and rapidly, the organization can address business opportunities quickly and at scale.

Not one to rule them all: Model your data for your industrial reality

To transform operations, industrial companies need to build 10s of data-driven solutions and then scale them across 100s production facilities.

Scalability helps enterprises break the Proof of Concept (PoC) purgatory cycle. While some companies have lighthouse sites, where technology allows different teams to work in harmony, achieving this performance state commonly takes 1-2 years. If an organization has 50 sites in total - do they have 50-100 years to achieve the same level of performance at each of those 50 sites?

To address this lack of speed, the industry needs an approach combining domain and industrial data expertise into a single product, enabling data reuse to develop many tailored solutions rapidly. Data modeling is a core component of turning siloed data into scalable solutions.

Physical, industrial systems are complex to represent, and there is no single representation that will work in all the different ways you need to consume the data. The solution to this complexity is standardizing on a set of data models that contain some of the same data but that allow you to tailor each model and add unique additional data.

For this, Cognite Data Fusion® provides a data modeling framework that allows different perspectives of the same data to be clearly described. We think about data modeling at **three** levels (Fig. 16):

- Source Data Model – data is liberated from source systems and made available in its original state.
- Domain Data Model – siloed data is unified through contextualization and structured into industry standards.
- Solution Data Model – data from the source and domain models that support specific solutions.

The different layers enable value creation holistically and on a per-project basis. While the source data model liberates data from a variety of source systems, it also maintains query ability in Cognite Data Fusion® through the same API interfaces. The domain data model allows for a higher level of entropy and the representation of evolving ontologies, while the solution data model is much more rigid while at the same time allowing for true scalability across two dimensions:

- **Scalability of one solution** is ensured through automated population of solution data model instances made possible by the contextualized relationships in your industrial knowledge graph, e.g., scaling a maintenance optimization solution across the entire asset portfolio.
- **Scalability across a portfolio of solutions** is enabled by immediate access to a wide range of data sources and the fact that application requirements to data are decoupled from the representation in the domain data model. This allows for use cases to be solved that require different levels of data granularity, e.g., plant-level maintenance optimization vs enterprise-level strategic planning.

As a result, enterprises can break free of the PoC purgatory cycle and focus on use case innovation that delivers in production at scale.

For a deeper dive into data modeling, check out out white paper:

Advancing Digital Twins with Data Modelling →

Data models work together in a digital twin

Cognite Data Fusion® further unlocks the potential of industrial digital twins-powered data transformation and automated contextualization services for developing and maintaining an AI-powered, open industrial knowledge graph. The industrial knowledge graph acts as the foundation for the data model of each twin and provides the point of access for data discovery and application development.

Industrial companies can enhance the overall understanding of their operations by creating relationships across OT, IT, and engineering data using contextualization pipelines to create an open industrial digital twin.

> An **industrial digital twin** is the aggregation of all possible data types and data sets, both historical and real-time, directly or indirectly related to a given physical asset or set of assets in an easily accessible, unified location. The collected data must be trusted and contextualized, linked in a way that mirrors the real world, and made consumable for a variety of use cases.

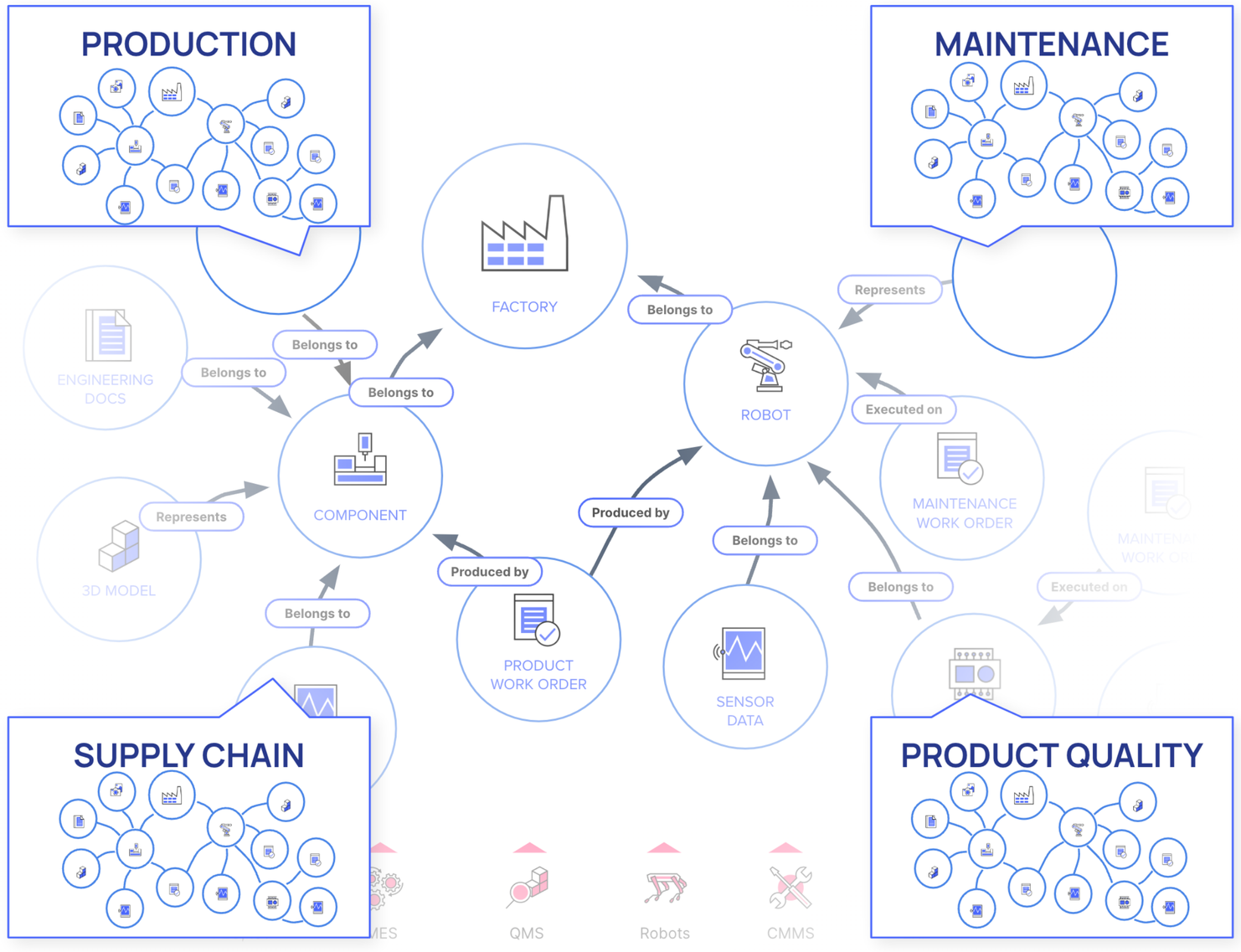
Digital twins must serve data in a way that aligns with the operational decision-making process. As a result, companies need not a single twin but multiple twins tailored to different decision types. For example, a digital twin for supply chain, one for various operating conditions, one for maintenance insights, one for visualization, one for simulation—and so on (Fig. 17).

Figure 17 shows that a digital twin isn't a monolith but an ecosystem. What is needed is not a single digital twin that perfectly encapsulates all aspects of the physical reality it mirrors but rather an evolving set of "digital siblings" who share a lot of the same 'DNA' (data, tools, and practices) but are built for a specific purpose, can evolve on their own, and provide value in isolation. To support that ecosystem, industrial companies need an efficient way of populating all the different kinds of digital twins with data in a scalable way.

Contextualized data generates significant insights for operators, increasing understanding and improving operations. With contextualized data, industrial companies find it easier to examine their assets across multiple levels, from individual sensors to complex models. Armed with virtual representations of real-world assets reflecting real-time data, operators can identify and prevent problems that have been present but invisible for decades.

As more people use and enrich the Open industrial Digital Twin, the quantity and variety of data in the digital twin become richer and more valuable than all other data sources combined.

▼ Figure 17: Industrial Digital Twins can capture the specific needs of the various business domains.

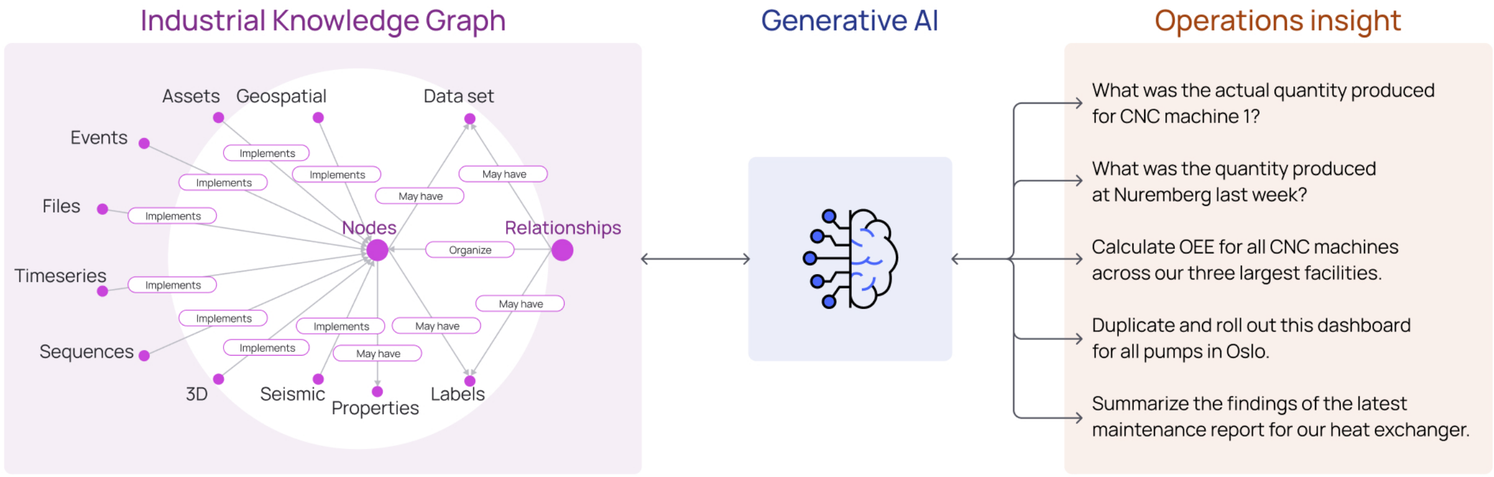


Industrial Generative AI requires data contextualization

Contextualized data generates immediate business value and significant time-savings in many industrial performance optimization applications, as well as across advanced analytics workstreams. Plus, access to contextualized data allows subject matter experts (SMEs) to become more confident and independent when making operational decisions or when working on the use cases with data scientists and data engineers. Contextualization helps solve the complexity of industrial data but once this is all done, the question is, how do we then use this data?

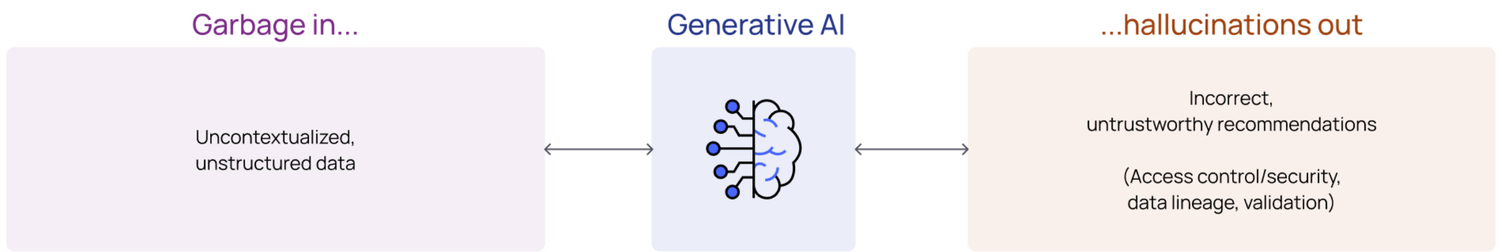
Our early exposure to Large Language Models (LLMs) through ChatGPT and Dall-E made the industry quickly realize the incredible generative capabilities of this latest advancement in deep learning AI. The question is how can we leverage this technology to accelerate contextualization and as a result, accelerate time to value, improve operations, and provide an even easier way for subject matter experts to consume and work with data. The 'generative' part of LLMs can be both a blessing and a curse. These models are trained to create new content (text, pictures, audio) following patterns they have learned during training. These generative results are, by design, plausible but not necessarily based on real facts in the absence of sufficient context. This is why we see such a tendency to hallucinate with responses that seem convincing but can be inaccurate (Fig. 18).

While generative AI can help to make your data "speak human," it doesn't speak the language of your industrial data on its own. A strong industrial data foundation is required to remove the risk of "hallucinations," or a response by an AI that is not correct or justified by its training data. Generative AI can improve data usage, but trustworthiness requires a robust data foundation with a powerful contextualization engine.



▲ Figure 19: Deterministic Industrial Knowledge Graph is a necessary prerequisite for generative AI technologies to to form an accurate understanding of an organization's industrial reality.

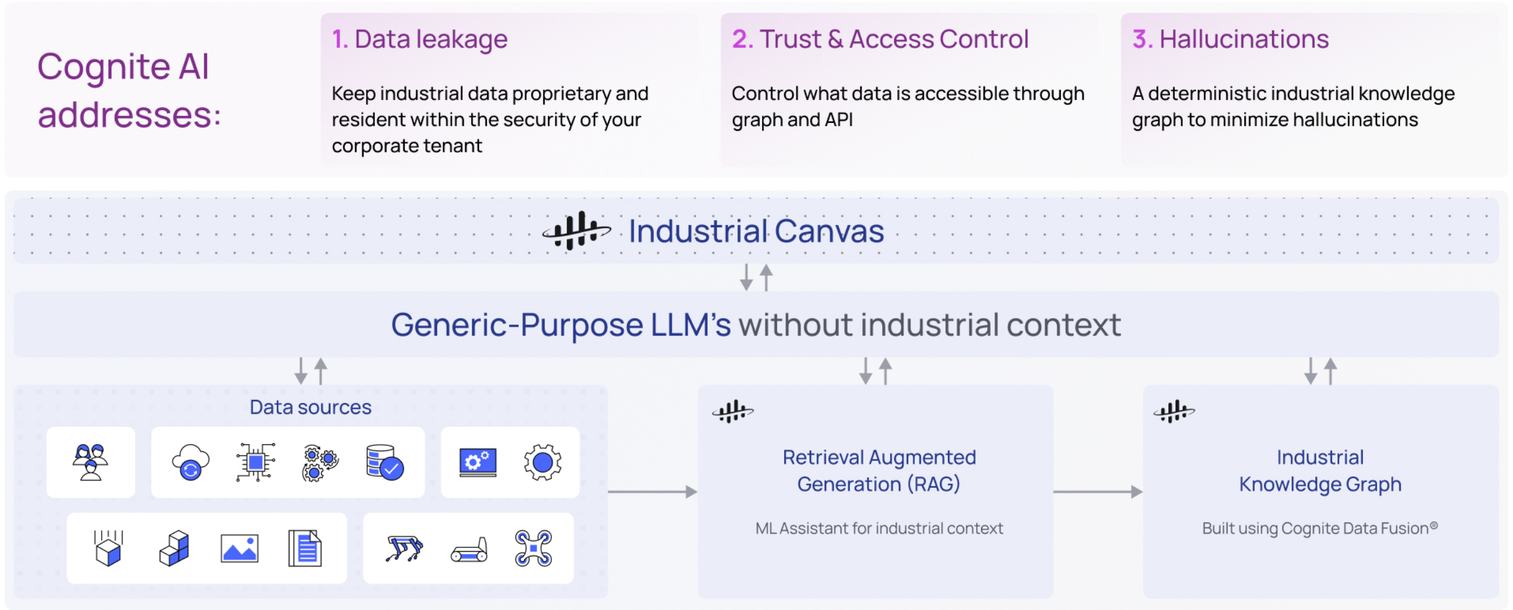
▼ Figure 18: Uncontextualized/unstructured data provides untrustworthy insights



This is why a deterministic industrial knowledge graph is a necessary prerequisite in any digital strategy that incorporates generative AI, serving as the reliable foundation for generative AI technologies to form an accurate understanding of your organization's industrial reality without hallucinations or data leakage (Fig. 19). The technologies are symbiotic. It is the combination of these two that allows your data to "speak human" and, more so, speak your specific organization's language.

However, there's another side to LLMs that is often overlooked amidst the generative excitement. The 'generative engine' of LLMs is powered by an incredibly advanced 'reasoning engine'. When harnessed and provided with rich contextualized industrial data, it is possible to use the 'reasoning engine' of LLMs to generate deterministic (fact-based) answers to natural language prompts (Fig. 20).

To capitalize on the 'reasoning engine' of LLMs, Cognite leverages Retrieval Augmented Generation (RAG), a design pattern used with LLMs to provide contextualized industrial data directly to the LLM as specific content to use when formulating a response. This approach utilizes the reasoning engine of LLMs to provide deterministic answers, based on the specific inputs we provide, rather than relying on the generative engine to create a probabilistic response based on existing public information. This serves as the reliable foundation for generative AI technologies to form an accurate understanding of your organization's industrial reality (do not expose your IP to a public generative AI solution).



▲ Figure 20: An innovative architecture that unifies Generative AI (LLMs such as GPT 3.5/4 and PaLM) with Cognite Data Fusion's specific Data Modeling and Retrieval-Augmented Generation (RAG) capabilities.

AI algorithms/models can also adapt and improve over time, continuously refining the contextualization process, enhancing the accuracy and relevance of the extracted insights and continually enriching the knowledge graph. While a deterministic industrial knowledge graph is a prerequisite in any digital strategy, enabling organizations to make faster and more informed decisions, particularly when integrating generative AI.

By utilizing the open APIs of major LLMs we can then leverage this trusted source of industrial context to create and store embeddings in a way that becomes searchable (semantically) and enables us (with minor prompt fine-tuning) to fully leverage the reasoning engine of LLMs to give us actionable insights.

By leveraging this pattern, we can keep industrial data proprietary and resident within the security of your corporate tenant. We can maintain and leverage the access controls required to maintain trust, security and audit requirements of large enterprises. Most importantly we can get deterministic answers to natural language prompts by explicitly providing the inputs that LLMs should use when formulating a response.

Use Cases for Industrial Generative AI

> **Generative AI will change how data consumers interact with data, enhancing self-service and potentially making some workflows human-less, but can it completely eliminate "human contextualization"?**

Whereas a process engineer might spend several hours performing 'human contextualization' (at an hourly rate of $140 or more) manually – again and again – contextualized industrial knowledge graphs provide the trusted data relationships that enable generative AI to accurately navigate and interpret data for operators without requiring data engineering or coding competencies.

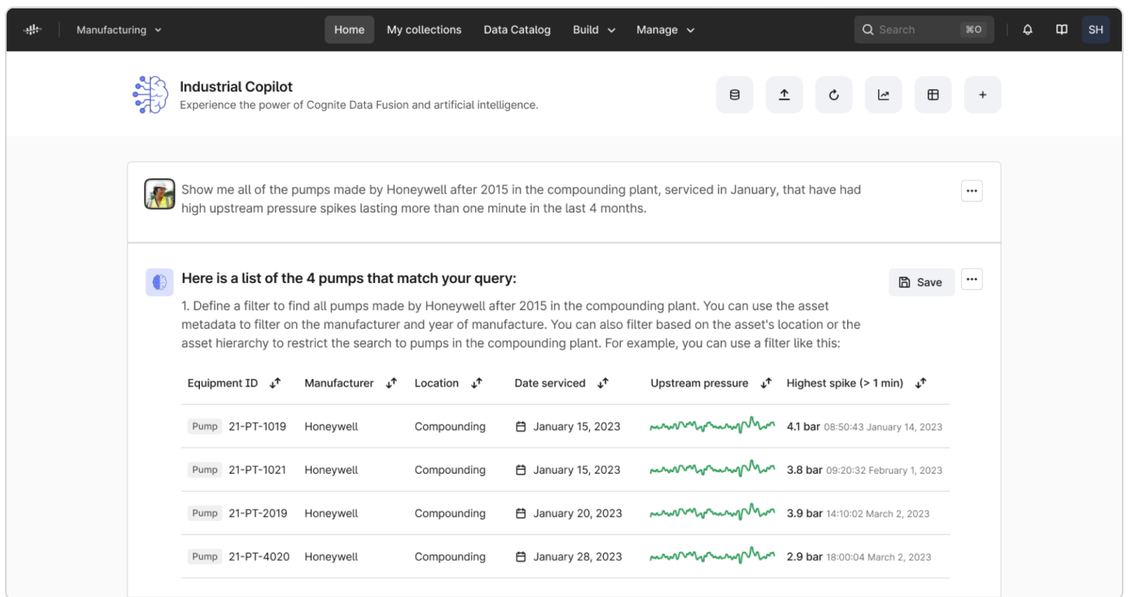
With generative AI-powered semantic search, what used to take your process engineers, maintenance workers, and data scientists hours of precious time will take only a few seconds.

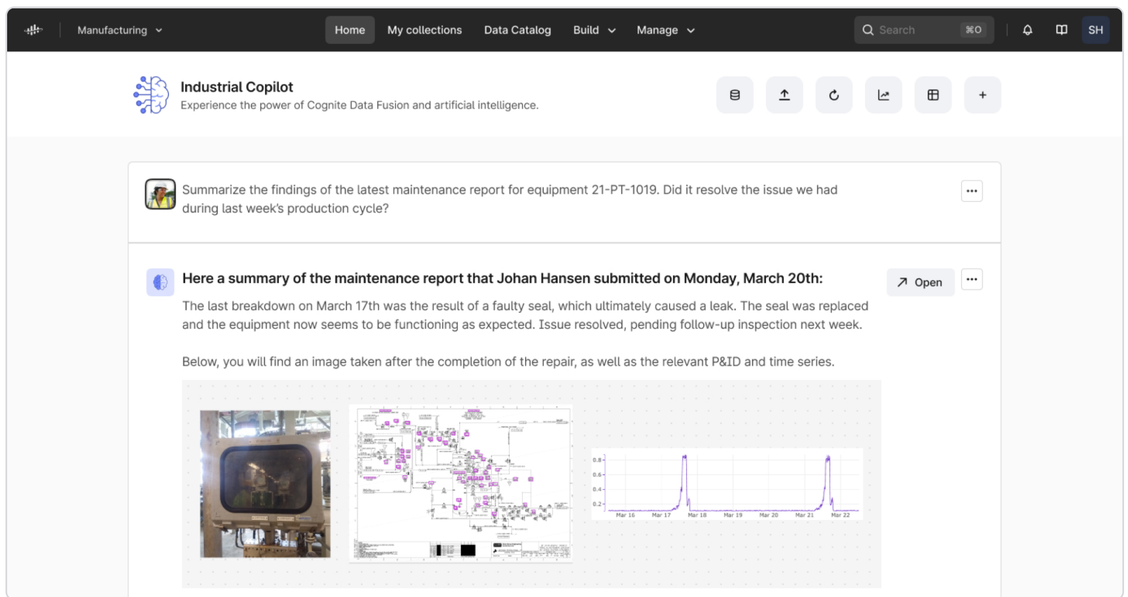
Take, for example, a Copilot approach. Because Large Language Models (LLMs) like ChatGPT understand and can generate sophisticated code in multiple languages (i.e., Python, JavaScript, etc.), we can prompt the LLM with a question about our industrial data, and it can interpret the question, write the relevant code using Cognite Data Fusion's APIs, and execute that code to return a response to the user (Fig. 21).

These copilot-based approaches leverage the power of natural language to understand and write code based on published API documentation and examples. This level of automation is impossible with data lakes or data warehouses where without a contextualized industrial knowledge graph, there are no API libraries that can be used as a reliable mechanism to access rich industrial data.

Cognite Data Fusion® can also provide contextualized data directly to the LLM API libraries available from OpenAI, LangChain, and others, allowing us to leverage the power of the LLM's natural language processing in conjunction with proprietary data. This database can include numerical representations (embeddings) of specific asset data, including

▼ Figure 21: Examples of Industrial Copilot search, enabling Google/ChatGPT-like search for industrial data





time series, work orders, simulation results, P&ID diagrams, and the relationships defined by the digital twin knowledge graph. Using these open APIs, we can then send a prompt to the LLM along with access to our proprietary embeddings database so that the LLM will formulate its response based on the relevant content extracted from our own proprietary knowledge graphs.

Cognite's Industrial Canvas makes data 'speak human'

For every one person that can 'speak code,' there are 100s who cannot.

Cognite uses generative AI to enhance data onboarding, complete with lineage, quality assurance, and governance, while a unique generative AI architecture enables deterministic responses from a native AI Copilot.

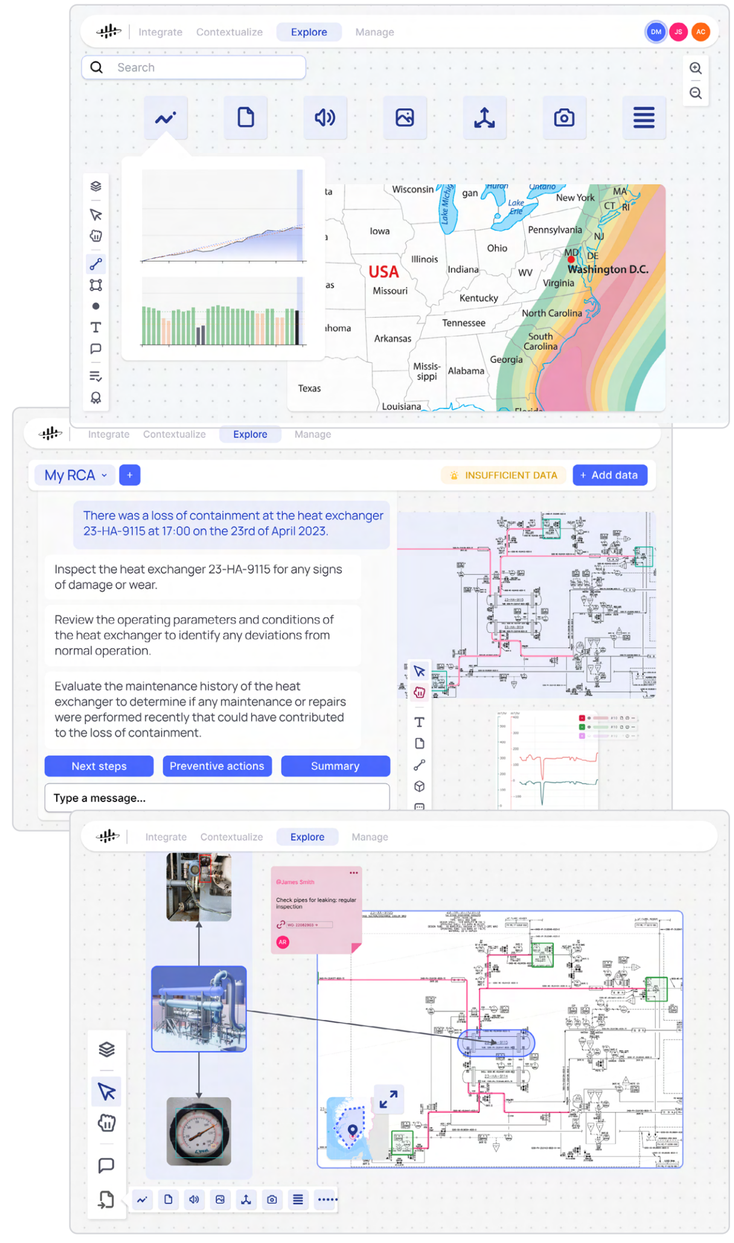
In that sense, Cognite Data Fusion's Industrial Canvas (Fig. 22) delivers the ultimate no-code experience within a free-form workspace to derive cross-data-source insights and drive high-quality production optimization, maintenance, safety, and sustainability decisions. Industrial Canvas is an intuitive, user-centric tool revolutionizing data exploration and visualization. It makes cross-data source insights available without relying on data scientists, data engineers, and software engineers to build specific use case solutions, enabling everyone, at every level of the organization, to spend less time searching for and aggregating data and more time driving high-quality business decisions.

In Industrial Canvas, users can view all data types in one place, including documents, engineering diagrams, sensor data, images, 3D models, and more. This is the data that matters to our industrial end users: process engineers, maintenance planners, reliability engineers, machine operators, technicians, and many more. Users can generate summaries of documents and diagrams, perform no-code calculations on time series data, and conduct root cause analysis of equipment with the guidance of the AI-copilot.

Industrial Canvas is a collaborative environment where users can choose to share workspaces, tag other users, and share insights as comments. This overcomes the single pane of glass solution which often over-promises as a framework as it is often difficult to collaborate and too rigid, preventing users from working with the data how they choose.

Industrial data Canvas is a perfect example of how contextualized data generates immediate business value and significant time-savings in many industrial performance optimization applications, as well as across advanced analytics workstreams. Plus, access to contextualized data allows subject matter experts (SMEs) to become more confident and independent when making operational decisions or when working on the use cases with data scientists and data engineers.

▼ Figure 22: Example of a data exploration and visualization workflow in Cognite Data Fusion's Industrial Canvas



In summary

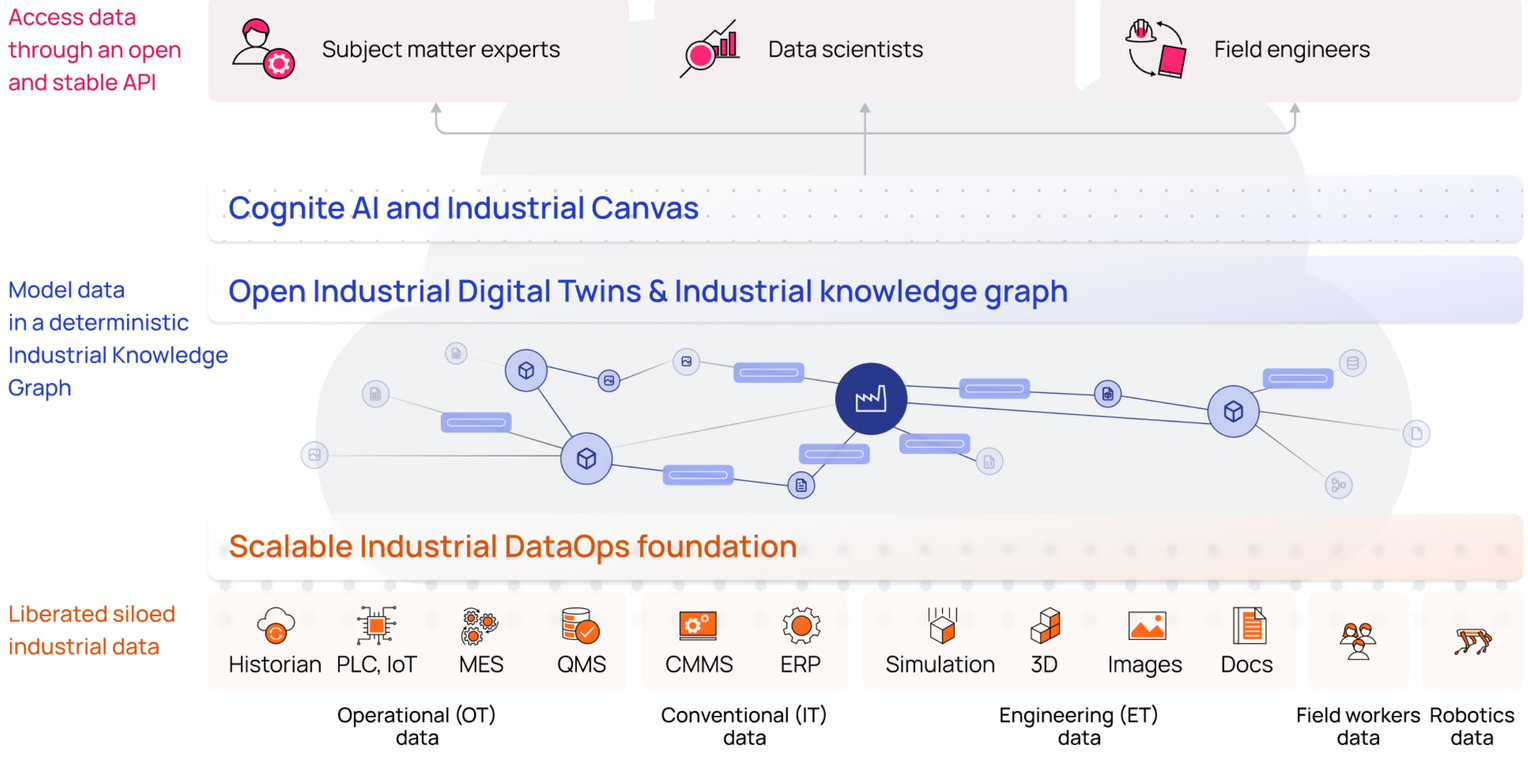
At Cognite, we're often asked which users benefit the most from contextualized industrial data. Should the focus be on a rich intuitive data exploration for subject matter experts (SMEs), production managers, business analysts and engineers — or providing a 'data as code' experience to data scientists with a preference for SDK experiences? We believe in **both**, a new way of serving all data consumers — data and analytics professionals as well as SMEs, business, and engineering professionals — with the same 'real-time contextualized data at your fingertips' experience (Fig. 23).

With continued loss of domain-specific knowledge caused by the aging workforce, various data types and source system complexity, and **painful user experiences** amongst the subject matter experts (SMEs), data contextualization is the only way to address the Achilles heel of all data and analytics solutions.

Data contextualization is for all to improve decision-making by enabling people to generate operational improvements, immediate business value, and significant time-savings in many industrial performance optimization applications, asset performance management, root cause analysis, and advanced analytics workstreams. Additionally, data contextualization is equally vital for scaling use cases and is critical to transform operations.

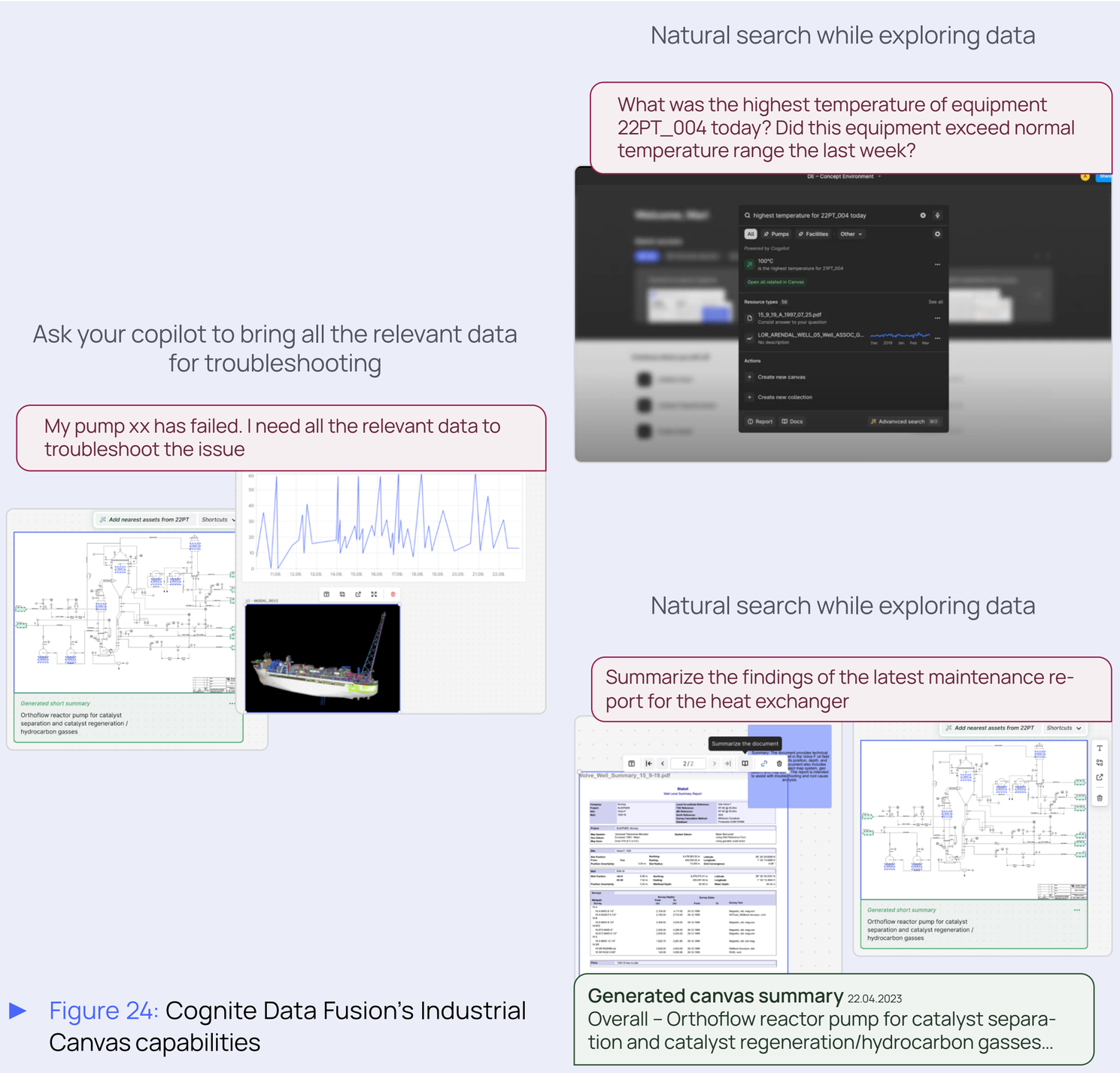
Cognite Data Fusion's Industrial Canvas is the latest release designed to deliver an ultimate no-code experience within a free-form workspace to derive cross-data-source insights. Industrial Canvas is a collaborative environment where users can share workspaces, tag other users, and share insights as comments. It is an intuitive, user-centric tool that revolutionizes data exploration and visualization and makes cross-data source insights available for process engineers, maintenance planners, reliability engineers, machine operators, technicians, and many more, to build specific use case workflows without relying on data scientists, data engineers, and software engineers.

▼ Figure 23: This is how Cognite Data Fusion®'s provides simple access to complex industrial data to subject matter experts, and continuously enriching this data via an industrial knowledge graph, which is available to everyone across the domains.



With the guidance of a generative AI Copilot, users can also generate summaries of documents and diagrams, perform no-code calculations on time series data, conduct a root cause analysis of equipment, and more. Time spent gathering and understanding data goes from hours in traditional tools to seconds. Now users can spend more time driving high-quality business decisions across production optimization, maintenance, safety, and sustainability (Fig. 24).

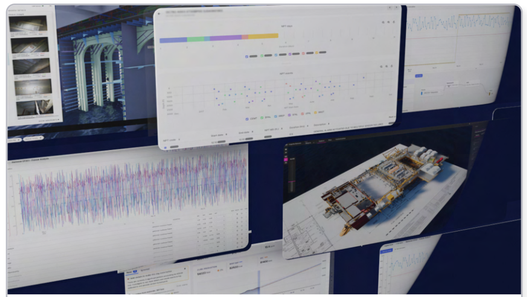
Creating this experience is predicated on Cognite Data Fusion's ability to contextualize data rapidly at scale with AI-powered services that eliminate tedious manual contextualization. The industrial knowledge graph is created through contextualization that serves as the connecting fabric between data modeling, digital twins, and all components within Industrial Canvas. Combining Cognite Data Fusion's industrial knowledge graph with a unique generative AI architecture, Cognite Data Fusion® delivers simple access to complex industrial data for all users. In their language. On their terms.



▶ Figure 24: Cognite Data Fusion's Industrial Canvas capabilities

Want to know more about our product?

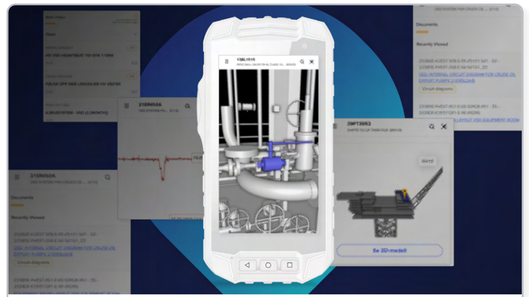
Explore more insights from Cognite



PRODUCT TOUR

Learn from Cognite customers and product managers how Cognite Data Fusion® simplifies and streamlines the data experience of a subject matter expert.

WATCH NOW →



CUSTOMER STORIES

Discover how Cognite Data Fusion® makes data more accessible and meaningful, driving insights that unlock opportunities in real-time, reduce costs, and improve the integrity and sustainability of your operations.

GO TO STORIES →

ANALYST REPORT

Customer interviews and financial analysis reveal an ROI of 400% and total benefits of $21.56M over three years for the Cognite Data Fusion® platform.

READ THE REPORT →

BLOG

Discover our rich catalog of industry insights and technology deep dives.

READ OUR NEWEST BLOGS →

COGNITE

COGNITE.AI →