Johnson&Johnson x Capgemini

Innovation
& Techday

Manufacturing optimization.

Leveraging AI and ML to maximize productivity and yields while predicting and reducing batch failures

April 2026

01 The challenges of the pharmaceutical production

Facing the challenges of the production – the example of bioproduction

Why this matters :

The API is twice the value of **Gold** | One batch is worth 22 **± 5M€, even** after 11 years | it last **~ 20 days** | "every **1/1000** is worth"

The fundamental PROBLEM

- Achieving near-nature efficiency in living, handcrafted production systems

The challenge is not production but its variability

The fundamental CAPABILTIES

- Analyze root causes
- Optimize yield

Obstacles

- Biological Process Variability
- Observability and Control
- Scale-Up and Scale-Out

System

- Huge complexity : multiples dimensions / kind of scales / hard to figure out
- Chaotic behavior : high volatility / low predictability
- Long term memory : effect of earlier stages, delayed correlation

Method

- Scientific rigor, evidence based
- Full mastery of knowledge & data
- Smart/ Safe Hybrid Artificial intelligence

02 Pharmaceutical manufacturing success stories

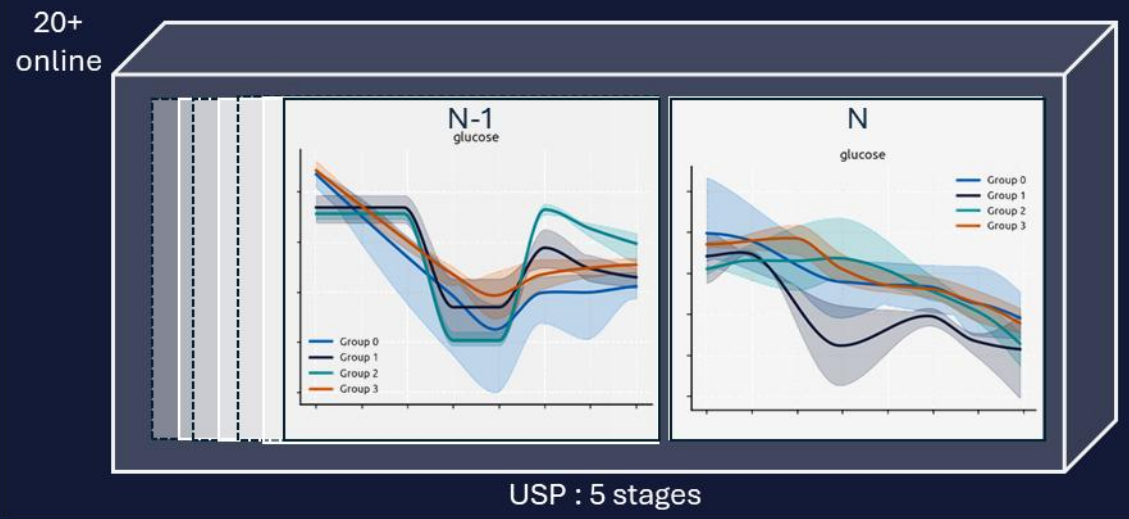
Golden tunnel : for variability reduction

Real-time batch trajectory monitoring and prediction : early abort in R&D / Increase yield in production

R&D → **PILOT PLANT** → **PRODUCTION**

- **Early stopping of poor R&D batches** speeds up Time-to-Market
- **Anticipate poorly performing batches** with 95% confidence
- Focusing on good runs **with +15% yields on average**
- **Correlate results with trajectory – "learn how to flight"**

- Batch variability is **above 20%**
- Models captures trends, not **intrinsic behaviors**
- Take good decision to **reduce low yield batches**
- Continuously monitor **process dispersion**



Introducing adaptative
process control
based on trajectory analysis

Staying GMP compliant

Success stories from the shopfloor (Brownfield case)

Digital App : Optimizing yield via shopfloor guidance - change & anomaly exploration toolkit for process engineers

Find valuable actionable insight – **within the qualified process**

Monitor gains

Onboard **data / IT teams upfront**

Synchronize with site operations

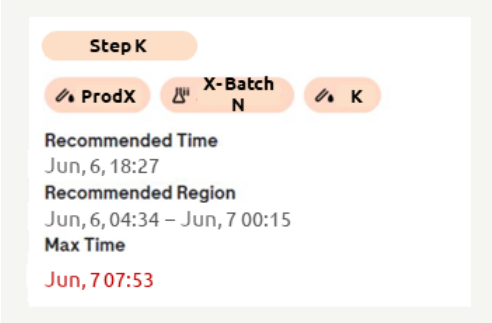
Re-architect impacted digital process

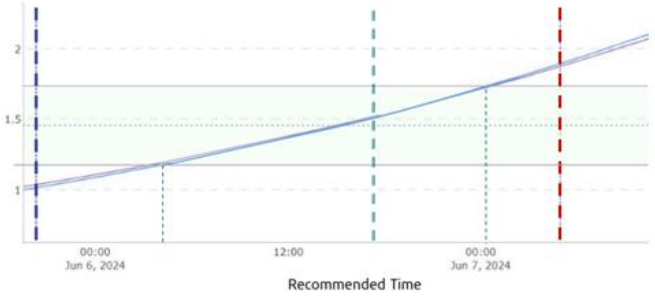
Develop and roll-out In preprod then prod.

Enable each function with seamless UX

Set-up Data / Model governance

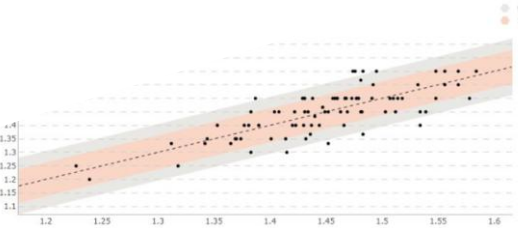
For shopfloor operator
- Clear guidance next action
- Optimal transfer time, adjusted on shifts





For Process Engineers
- Predictive tools
- Extensive statistical analysis
- Real-time anomaly detection
- Alerting system with notification

For Data scientists & Data Engineers
- Full MLOps pipelines (Drifts,…)
- Traceability and logging



Roche Digital Twin for Bioprocess Yield Optimization

Context:

Roche, a global pioneer in pharmaceuticals and diagnostics, is revolutionizing its bioproduction processes by implementing an advanced Machine Learning solution. By optimizing manufacturing processes, the company aims to increase product yield, reduce costs, gain valuable insights through data analysis and maintain consistent quality.

Challenges/Pain points:

- Difficulty in achieving optimal product yield, with > 10% variability
- Lack of data-driven informed decision to adjust process actions
- Very little knowledge extracted from MES data actionable for process engineers

Gain points:

- Estimated +>1 % improvement of the yields over the first batches gained by optimized batch transfer time
- Deep-dive capabilities for anomaly detection of both online/offline process measured parameters
- Scaled across 2 sites – extension on-going.

Approach & Solution:

Design and develop a scalable bioproduction monitoring pipeline on Dataiku, using Machine Learning to:

- Predict yield and optimize batch transfer timing
- Implement MLOps solutions for anomaly and drift detection
- Create a WebApp for visualization and data analysis

BENEFITS DELIVERED

+ > 1 % Increased product yield

Real-time Shopfloor Decision making

Anomaly detection to process engineer

03 Prerequisites for leveraging data and AI

Prerequisite : Create a solid actionable *knowledge* foundation

Leadership decisions shape production for years — poor decisions cannot be rolled back

Secure path toward valuable use case

Technologies evolve, but the path to confident, auditable insight and GMP-compliant decisions remains invariant

Design **business-driven data models** grounded in manufacturing processes — the challenge is a process-engineering problem, not a data-engineering one

1

Define a **transversal functional scope** across the full value chain,
— and validate ROI accessibility consistently across scales. – Disconnections sink the value

2

Provide **direct, technology-agnostic access for end users**, with validated analytical insights and evidence-based recommendations

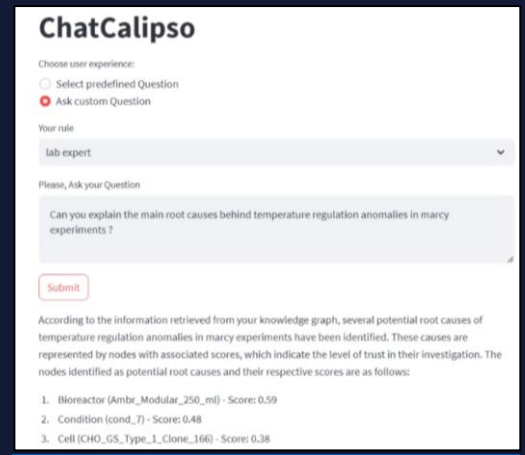
3

Establish a **robust organization and governance model** that provides the **right decision** frameworks and methods for the **right tools**.

4

Quality by Design frameworks integrating process models, semantic context, and governed toolchains to enable controlled process evolution

5



A replicable journey, not a sprint

04 Unifying industrial data for enterprise value

What problems are we trying to solve?

How to improve quality, compliance and operational efficiency

From...

- ☒ Limited operational visibility
- ☒ Limited performance insight leading to over maintenance
- ☒ Time and labour-intensive analysis
- ☒ Lack of easy access to operational know-how

Siloed Systems

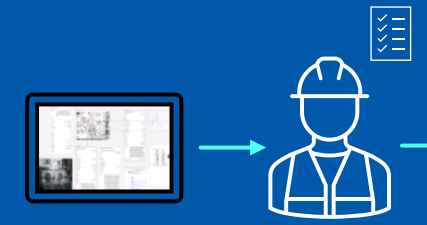
- Sensor Data
- Work Orders
- Maintenance Records
- Equipment Manuals
- Standard Operating Procedures
- Experienced Engineers

Catch & resolve issues based on experience

...To

- ☑ Holistic operational visibility leading to streamlined operations
- ☑ Auto generated insights
- ☑ Easy access to knowledge & expertise
- ☑ Efficiency gains and productivity

- Sensor Data
- Work Orders
- Maintenance Records
- Equipment Manuals
- Standard Operating Procedures
- Experienced Engineers

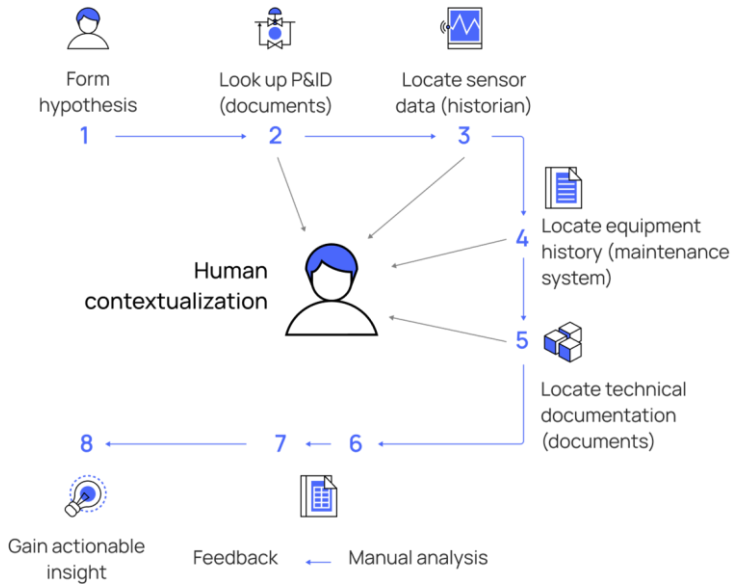


Leverage Canvas and AI insights

Cognite Data Fusion is one of the solutions

From this …

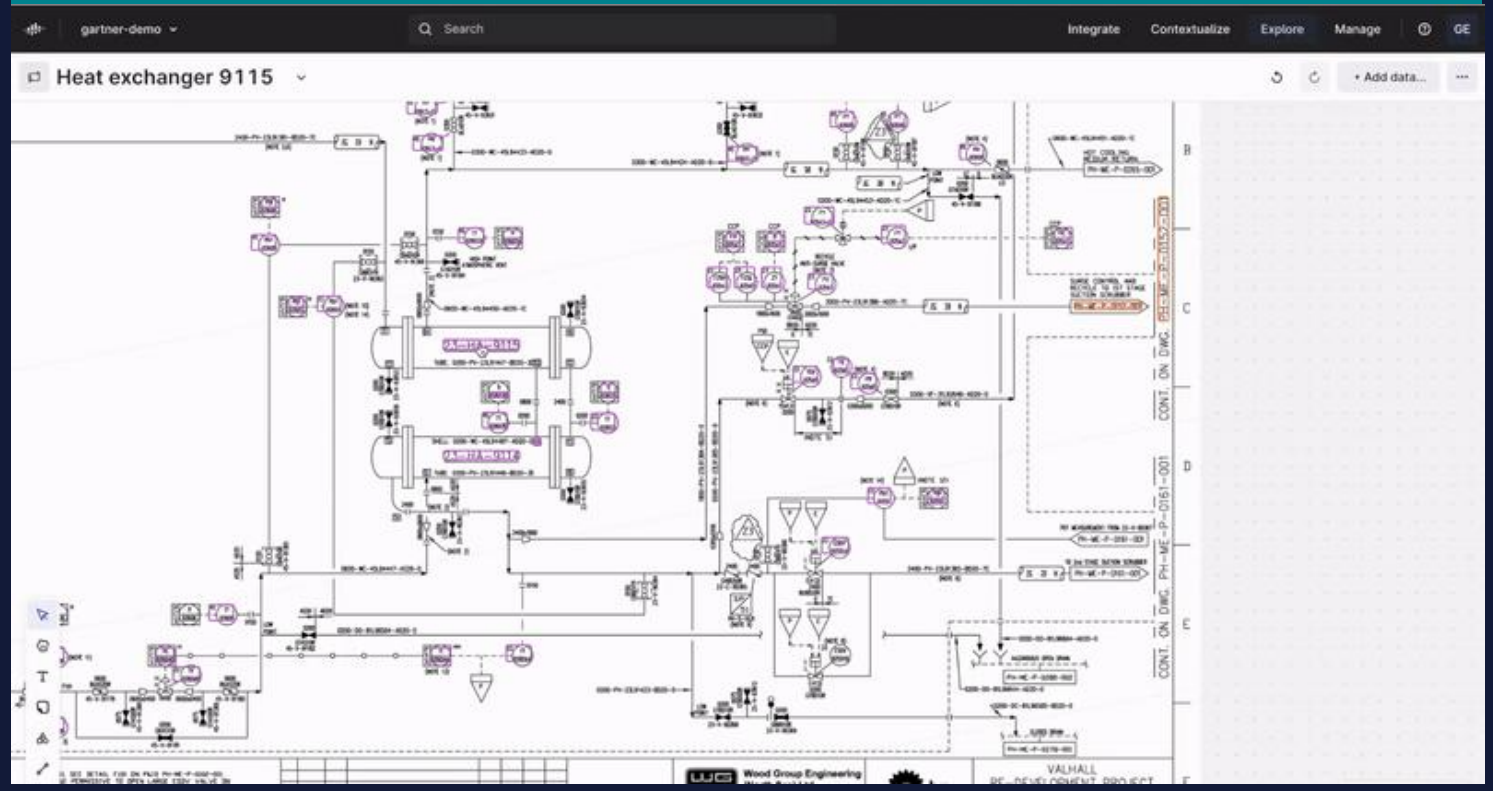
Manual, cumbersome industrial data workflow



No simple access to complex industrial data and insights

… to THIS

Single workspace for data & analytics powered by AI



Explore
Access and add any type of data in one compostable environment + Copilot powered research

Collaborate
Annotate, tag users, share insights with seamless integration into no-code analytics

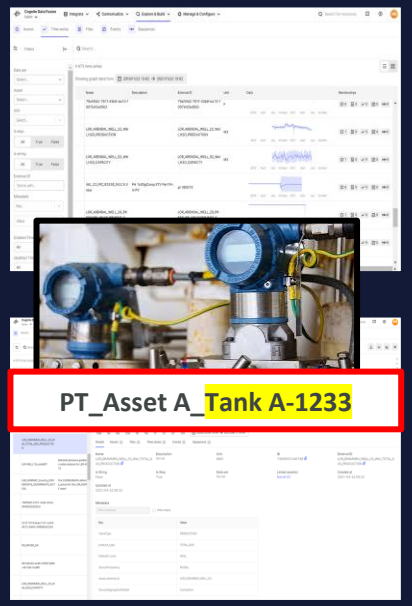
Create
Summarize documents and write complex queries using natural language

Industrial data is diverse and complex ... and inconsistent

Assets

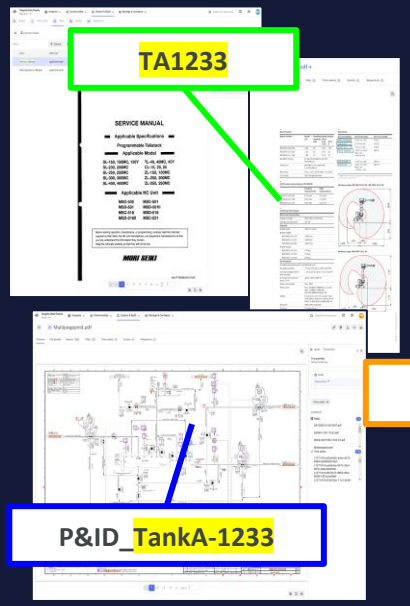
E.g. Equipment hierarchy from engineering tools or CMMS systems. Possibility to represent multiple hierarchies (e.g. asset, product).

Time series



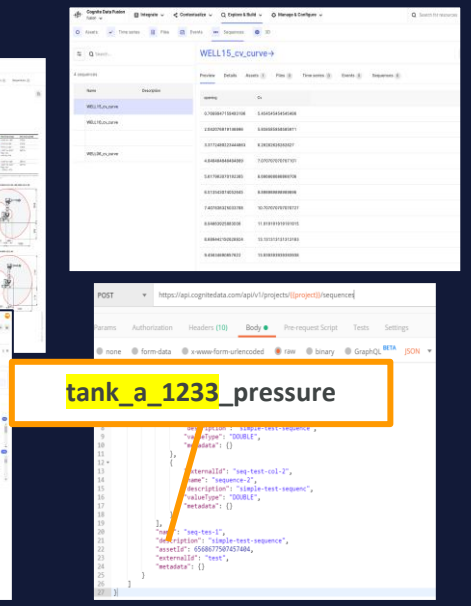
E.g., real-time readings of temperature, pressure, speed, number of good parts produced, number of scraps and other KPIs.

Files



E.g., 2D CAD and scanned drawings, service manual, engineering diagrams, such as P&IDs and isometric drawings, and Bill of Materials.

Sequences

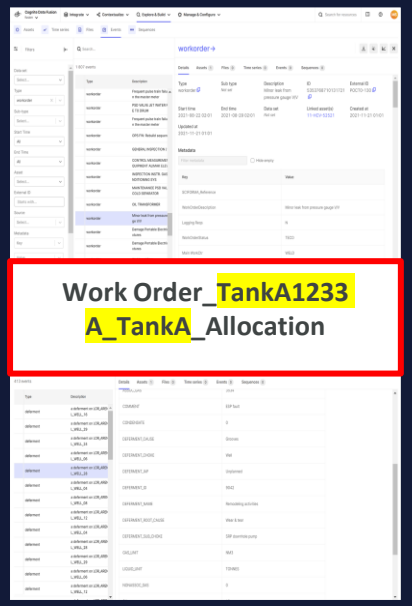


E.g., operating curves, tabular data, multivariate data and LIMS data.

Visual data

E.g., 3D CAD model of entire plants or components, images of equipment, laser scans/point clouds and 360 images.

Events

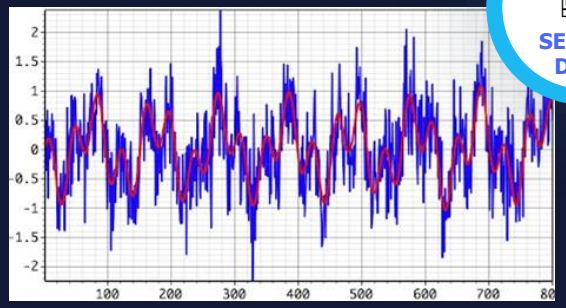


E.g., real-time system failures and alarms, identified abnormalities, work orders, conducted inspections and production orders.

Time taken for contextualisation	35 mins
Contextualisation %	100%
Accuracy %	97.7%

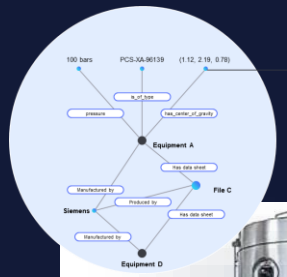
Data Contextualisation accelerates root cause identification

Time Series



SENSOR DATA

Prebuilt Extractors i.e. OSI pi, MQTT



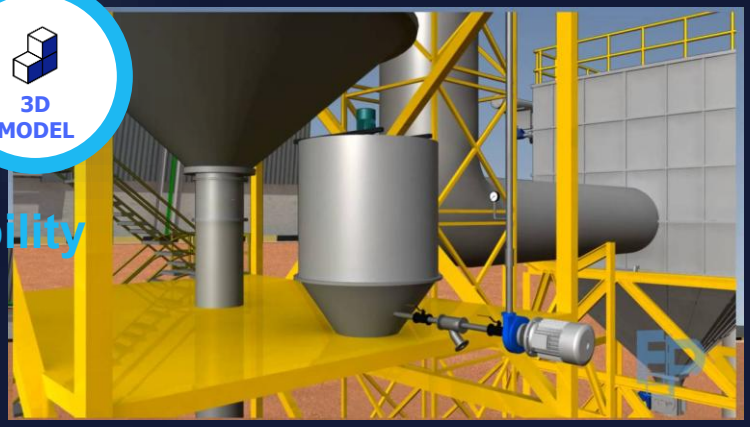
Knowledge Graph

ASSET

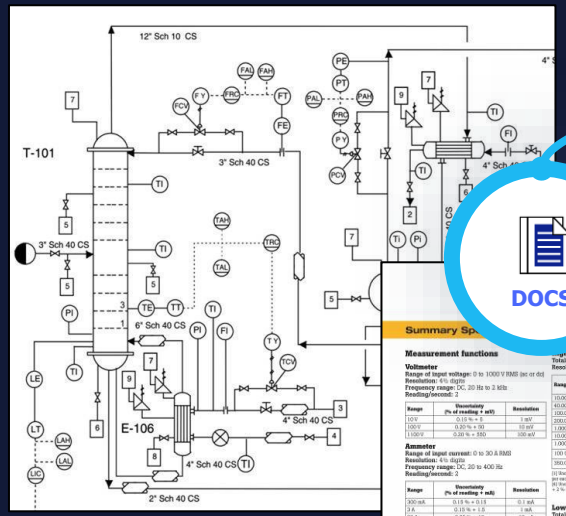
PT-Tank-A1233

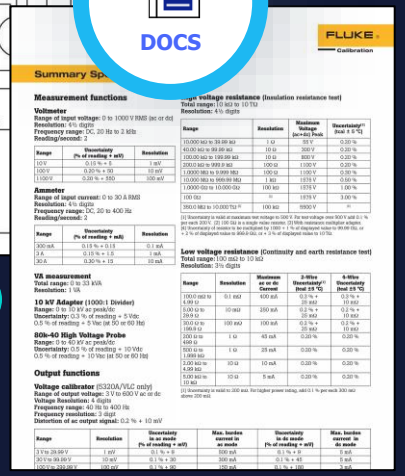
3D MODEL

3D dwg compatibility i.e. nwd



3D / 3D Point Cloud





Engineering Documents

OCR adaptor

DOCS

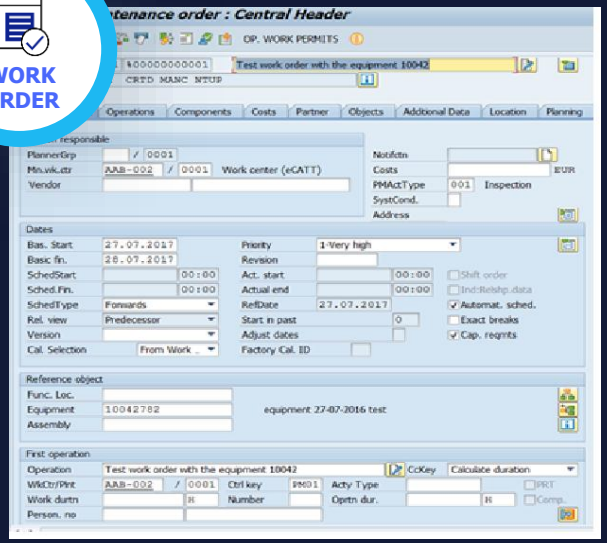
CV connectors

IMAGE

Photo / IP Camera

SAP templates

WORK ORDER



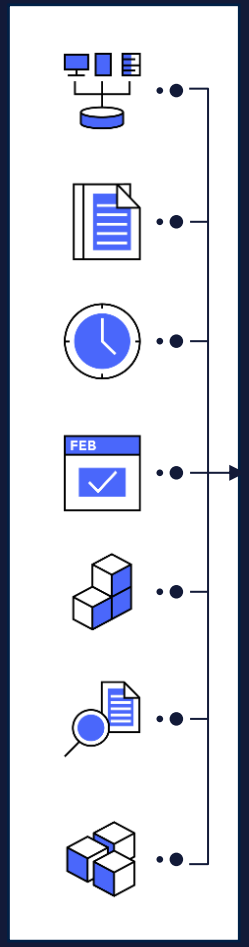
Maint. Work Order

Contextualization → Real-time Industrial Knowledge Graph

Data sources

Contextualization services

Industrial Knowledge Graph



Entity matching

Automatic and expert-guided resource mapping (assets, time series, events) to appropriate counterparts creating belongs-to relationships.

Contextualize visuals

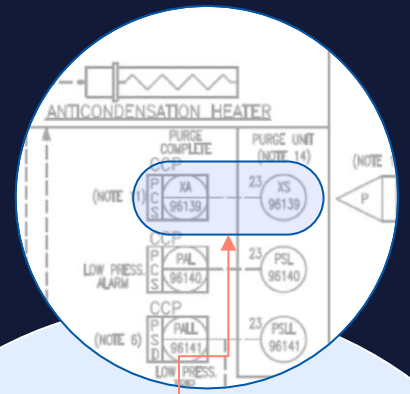
Contextualize imagery data, e.g.,automatic detection of assets, tags, sensor gauges, people, rust, birds, etc. in images.

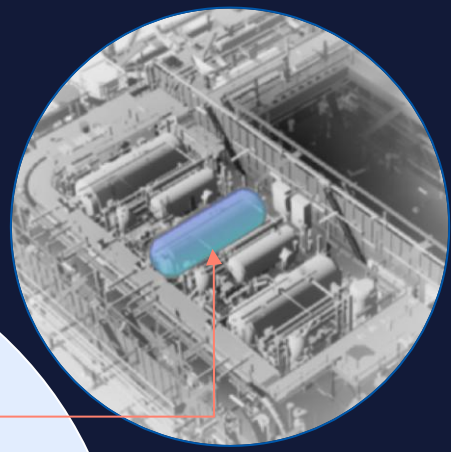
Document parsing

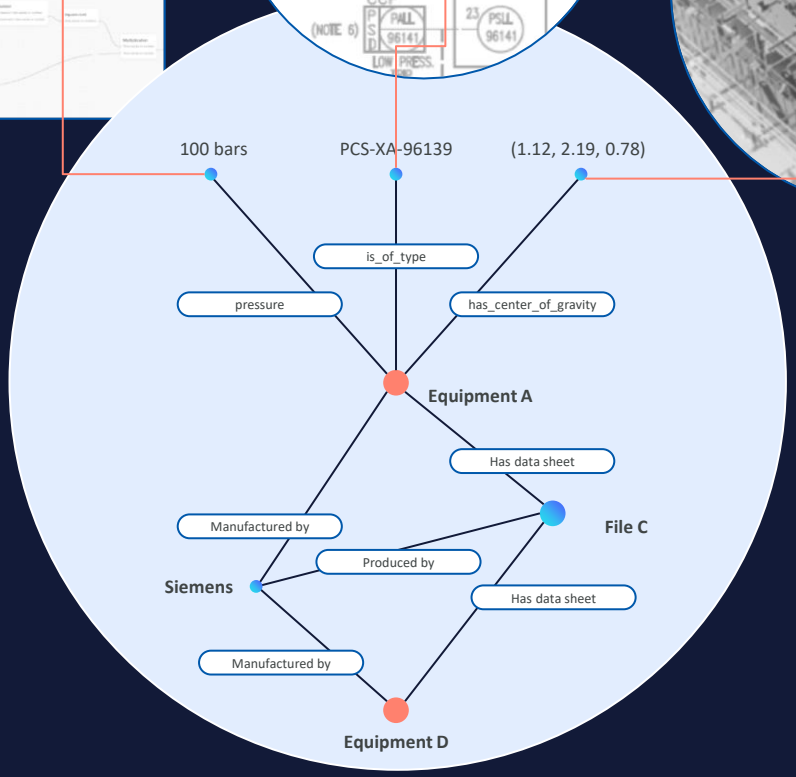
Automatically parse engineering diagrams (e.g., P&IDs) and documents, augment them with your domain knowledge to extract information, classify and make them interactive.

3D contextualization

Contextualize 3D models and enable applications with interactive exploration of e.g., assets, time series, and events.







CASE STUDY

Unlocking data for autonomous manufacturing in life sciences

Data Contextualisation for asset reliability uptime:

Across industries, companies are racing to attain ‘autonomous manufacturing’ - the ability to operate factories and machines remotely, with automated adjustments to optimise output

Challenges & Business Drivers:

- Industrial data is often siloed and owned by different entities in different format
- A single interface is needed to break these silos and make sure that the information is easily accessible to all
- Additionally, different owners across the lifecycle of an asset define the data based on their own empirical definitions that are not written or widely standardised
- A standardization of this data along the lifecycle is necessary to enhance digital continuity

Delivered gain points:

- 70% less time searching data
- 20% increased asset lifetime
- 80% reduced risk of human error
- Increase capacity utilization by 10%
- 15+ Capgemini certified staff
- 2 PoC Delivered, 1 pilot underway, global scaleout progressing

Approach & Solution:

- Cognite Data Fusion (CDF) is an industrial DataOps platform offered by Cognite
- AI-powered contextualization to create relationship between siloed sources and create industrial knowledge graph
- Automatically populates data models
- Summarize documents and write complex queries

BENEFITS DELIVERED

>£3M savings Increased machine uptime per year per site

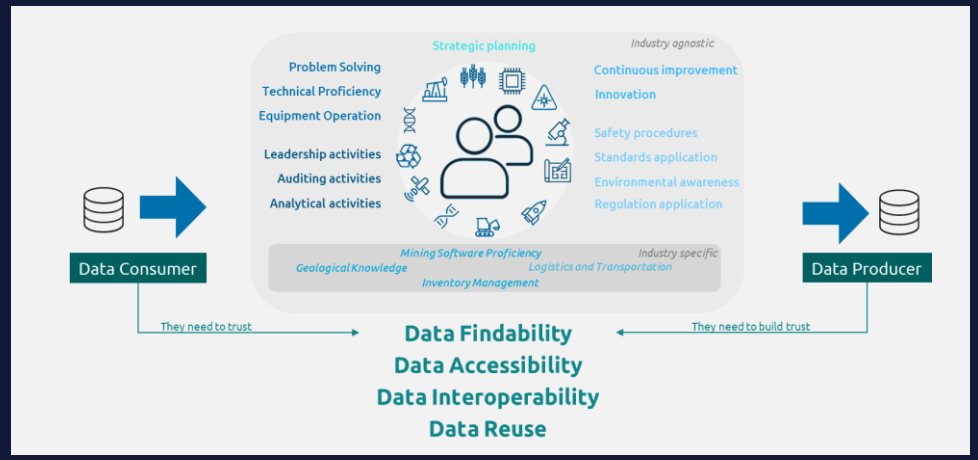
10%-20% Reduction in Energy cost

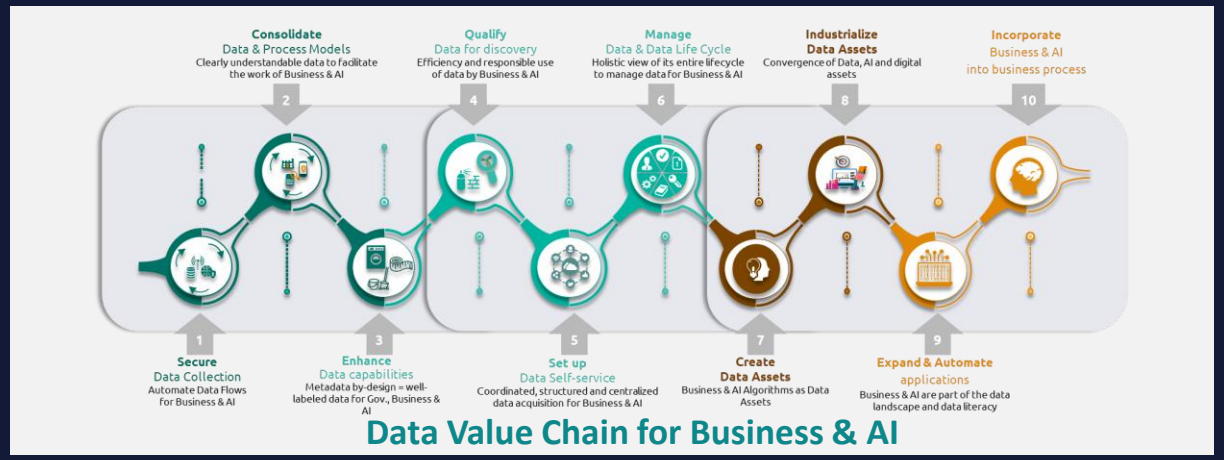
15% reduced maintenance cost

05 Key lessons learnt

Key Lessons learnt

1. **Value driven use case focus :**
   - Improve yield and performance within existing qualified processes, with clear safety and business
2. **Trusted, auditable data foundation :**
   - One shared data language, clear ownership and full traceability to support confident, compliant decisions





3. **Process led operational control**
   - Manage events and change across data models and operations with clear ownership and SLA's using automation, vision and robotics where they add measurable value

Experts

Patrick Chareyre
Director of AI for Engineering

patrick.chareyre@capgemini.com

Let’s connect!

Graham Upton
Head of Technology & Innovation

graham.upton@capgemini.com

Let’s connect!

Johnson&Johnson x Capgemini

Thank you.

Make it real.