CONTINUA PROCESS SYSTEMS

Date: July 24, 2025

WHITE PAPER

AI-Enabled Data Integration and Quality Management for Agile Pharmaceutical Manufacturing

Authors:

Paul Brodbeck
Chief Technologist

Anthony DeBiase
Director, Digital PD

Brian Sauerborn
Eng. Manager, DX

Continua Process Systems has developed an AI-enabled data integration platform that organizes industrial data into a central hub, contextualizes it with industry standards-based knowledge graphs, and transforms it into actionable intelligence. This platform enables and supports advanced technology for agile and distributed manufacturing of critical medicines with the goal of facilitating US onshoring. This data centric solution enables a lean quality management system to provide an end-to-end platform that generates a digital regulatory framework for the manufacturing process to align with the FDA's KASA and PQ/CMC initiatives for electronic regulatory submission, filing, and approval. The end goal is to expedite the end-to-end time to design, build, and commission an agile plant utilizing advanced manufacturing with streamlined CMC regulatory support to cost efficiently enable the onshoring of pharmaceutical manufacturing.

Continua has over 12 years of experience in advanced technology including agile continuous manufacturing, process analytical technology (PAT), digital transformation, automation, and machine learning. Continua is currently working with a leading university to deploy this data platform for Proof-of-Concept. Continua is affiliated with Control Associates, Inc., a recognized Emerson Process Management impact partner, providing DeltaV™ distributed control systems and systems integration services.

As most pharmaceuticals are made outside of the US, there is a need to onshore the manufacture of APIs (Active Pharmaceutical Ingredients), KSMs (Key Starting Materials),

and FDFs (Finished Dosage Form) to secure the domestic supply chain. There are many ongoing initiatives to enable onshoring, including the reduction of manufacturing costs due to higher labor costs and more stringent environmental regulations. One way of reducing costs is by using advanced technologies such as continuous manufacturing that support smaller, more agile distributed plants. In this paper we discuss technologies that support this approach by utilizing data centric manufacturing. Along with a digital quality system, data centralization can be a valuable component to streamline the end-to-end time to design, build, commission, and receive regulatory approval.

The FDA is currently modernizing its regulatory approval process from document centric to a more digital forward, data centric approach allowing them to shorten the review cycles. This solution will generate a digital regulatory framework that can be electronically transmitted to the FDA to align with the KASA and PQ/CMC initiatives for electronic regulatory submission, filing, and approval. KASA, Knowledge-aided Assessment & Structured Application, is a data centric framework designed to modernize the FDA's quality assessment process for drug applications by promoting structured, knowledge-driven evaluations. PQ/CMC, which stands for Pharmaceutical Quality/Chemistry, Manufacturing, and Controls, is a modernization initiative for data standards and processes used by the FDA for reviewing and approving drug applications. The FDA has issued multiple Federal Register Notices (FRNs) seeking public input on structured CMC data and PQ/CMC draft specifications. As new sections of Module 3 eCTD (Electronic Common Technical Document) are structured and standardized, FDA publishes these through FRNs to solicit industry comments and feedback. Module 3 of the Electronic Common Technical Document (eCTD) is the designated section for submitting Chemistry, Manufacturing, and Controls (CMC) information. As part of its PQ/CMC initiative, the FDA is working to structure this module by developing standardized data elements and controlled vocabularies to enable more consistent and efficient regulatory review. The latest draft was issued in September 2024.

The proposed technology solution consists of two integrated components: an AI-enabled data integration platform and a lean digital quality management system.

Technical Solution – Part 1: AI-Enabled Data Integration Platform

The first component of the proposed technology solution is an AI-enabled data integration platform, which encompasses several key functions, including:

1 – Unified Namespace (UNS)

Collect, centralize, and standardize data from multiple siloed sources into a Unified Namespace (UNS). A UNS serves as a central data hub that enables publish/subscribe

communication across the enterprise and functions as the *single source of truth* for information exchange between systems, devices, and applications. It is a real-time, taxonomy-based architecture that contextualizes and organizes data within a hierarchical, tree-like structure, facilitating scalable and structured access to industrial information. A well-implemented UNS mitigates many data analytics problems such as insufficient security access, multiple/disparate databases, multiple protocols, and lack of contextualization. Lack of contextualization is a typical problem in big data applications, for instance, where you might have data but not know exactly what asset the data is related to.

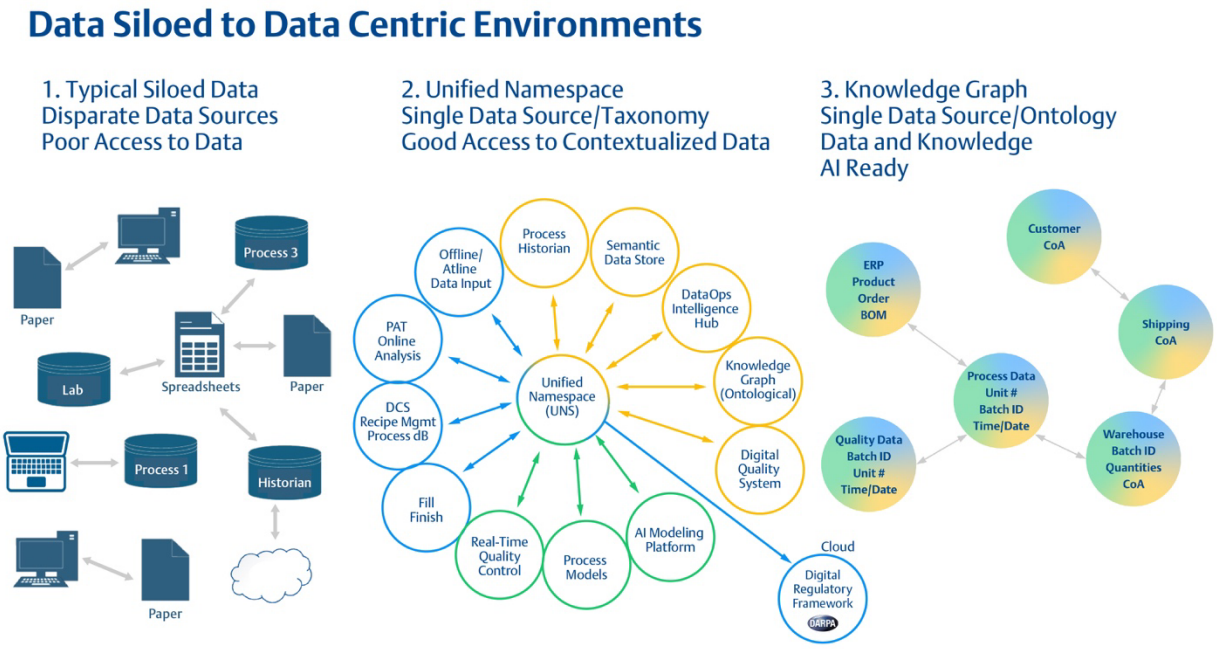

Figure 1 – Siloed data to UNS to Knowledge Graph

2 – UNS to Knowledge Graph

Data is further processed from the UNS into a knowledge graph to enhance the data for more context-rich, relationship-based reasoning and analytics.

A UNS organizes data structurally but may lack semantic relationships needed for higher-order analytics, such as root cause analysis which are often inferred by domain subject matter experts (SMEs). Within the UNS, data is digitally captured and made centrally accessible through a unified data architecture, often using a common communication protocol such as MQTT to enable standardized, real-time interoperability across systems. In addition to experienced data analysts, SMEs are often required to provide contextual knowledge and support complex data interpretation to enable the development of artificial intelligence (AI) and machine learning (ML) applications. To streamline the development of

AI and ML models SME's knowledge can be systematically embedded within a knowledge graph, enabling non-experts to perform data analytics and reasoning by leveraging its structured relationships and contextual semantics. A knowledge graph in this solution is used to enable AI models to function with minimal assistance from domain experts.

Figure 2 below depicts the effort required to track a production order from the time the order is issued to the plant to the time the product is sent to the customer.

For the siloed data it can be challenging and time consuming. The various databases are often not linked, and it is typical that data exists only in spreadsheets, or at worst on paper. Tracking an order to the plant, through production, the lab, warehouse, and shipment to customer is often done manually. Even though data in the UNS is centrally located and standardized, tracking the data from a production order through the system may still take several SMEs to understand the context from ERP, to plant, lab, warehouse, and shipment. While the data is being captured, it just may take people with domain expertise to focus on the targeted data and how the relationships between data flows. For the knowledge graph, the relationships (edges) can be traversed from node-to-node through the database to trace and relate a customer production order to the plant batch ID, process data, batch data, quality/lab data, warehousing, shipments, to final customer CoA.

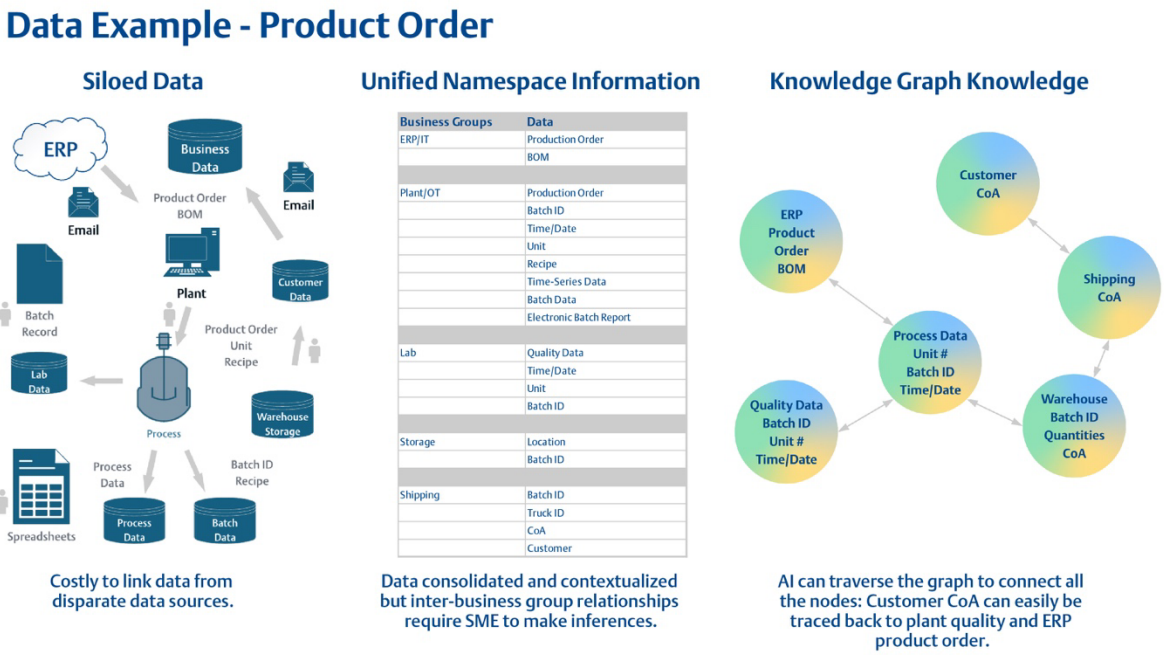

Figure 2 – Tracking a product order in 3 different modes

3 – *Knowledge Graph*

Knowledge graphs support reasoning through their semantic structure, enabling AI systems to perform logic-based analysis.

Artificial intelligence (AI) algorithms can be more effective when combined with knowledge graphs in applications that require contextual awareness, semantic integration, and relationship-driven inference. Knowledge graphs support reasoning by organizing data in a machine-readable, semantically structured format that enables logic-based AI analysis. Their graph-based node and edge structure is inherently more efficient for multi-hop traversals than traditional relational databases, enhancing the performance of AI systems that rely on complex relationship inference.

A knowledge graph is a data structure that represents information as a network of entities (nodes) and the relationships (edges) between them. It captures real-world concepts and their interconnections in a machine-readable, semantically enriched format. Each node typically represents a distinct entity (e.g., product, material, person, batch), and each edge represents a defined relationship (e.g., *produced by*, *tested at*, *belongs to*). Knowledge graphs are widely applied in domains that require contextual understanding, semantic reasoning, and relationship-based querying—such as pharmaceutical manufacturing, supply chain management, and regulatory intelligence.

To create a knowledge graph, an ontology defining the schema is required. The ontology provides the formal structure, rules, and controlled vocabulary that govern how entities and relationships are defined, organized, and interpreted within the graph. This semantic framework—embedded within the knowledge graph—enables machine-readable representation of domain knowledge and supports AI-driven reasoning, querying, and analytics.

Developing the ontology is a critical step in building an effective knowledge graph. Continua leverages the expertise of multiple industry consultants to define and structure the ontology tailored to the specific requirements of this application.

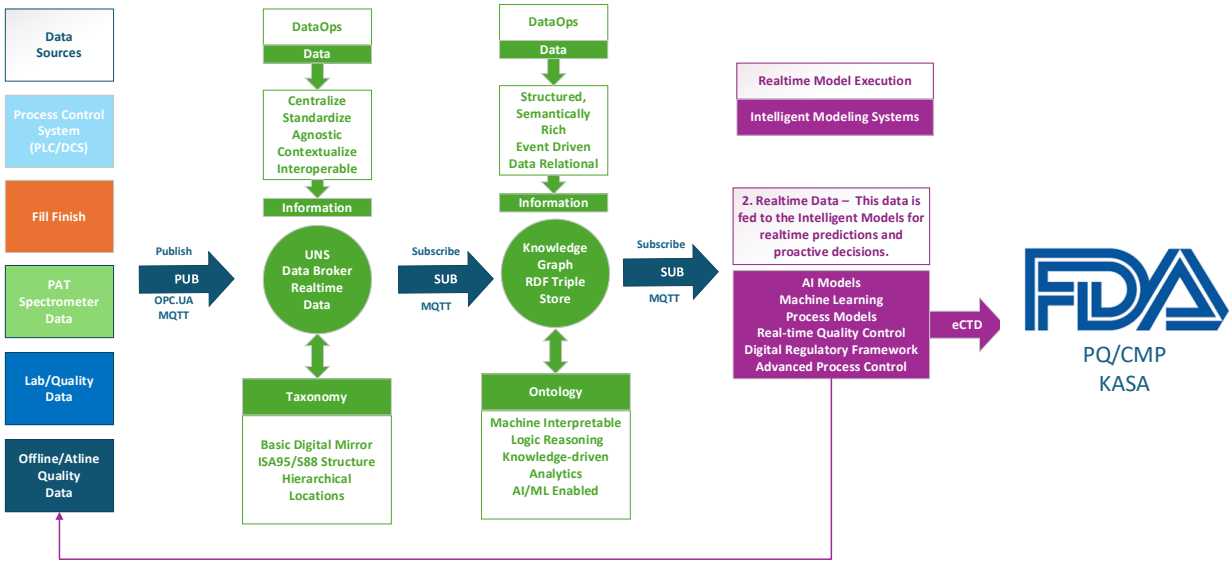

Figure 3 – Data flow from siloed, vendor locked sources to UNS, to knowledge graph, to AI, to FDA.

Technical Solution Part 2 – Lean/Digital Quality Management System

The second part of the technology solution is the lean digital quality management system. This system is built upon the AI-Enabled Data Integration Platform Technical Solution discussed in Part 1. Without the ability to qualify a pharmaceutical plant and manufacturing process in a reasonable amount of time, this technology has less impact. In addition to qualifying the system quickly it is also important to be able to quickly package the data for the FDA submittal. This is where the lean digital quality system and the digital regulatory framework need to be tightly integrated. The data from digital quality system flows directly into the data pipeline and is assisted by AI to generate the digital regulatory framework. This solution is designed to support compliance with 21 CFR Part 11 and GAMP5 principles and is structured to maintain real-time audit readiness for regulatory inspection. The main components of this system are as follows:

Lean Quality Management

- Model Process Design
- End-to-End Process Digital Map
- Integrated system (ERP, MES, LIMS, QMS) to create a single source of truth.
- Waste Elimination through Automation
- Streamline documentation and validation using digital signatures, templates, and workflows.
- Digitally enforce standard operating procedures (SOPs) while allowing version control and contextual adaptation.
- Data-Driven Risk Management

- Use predictive analytics and machine learning to proactively identify and mitigate quality risks.
- Implement risk-based decision-making across the quality lifecycle.
- Use data analytics to validate effectiveness of improvements.
- Integrated Compliance – align digital processes with regulatory requirements. Maintain audit readiness through traceable, time-stamped, and validated digital records.
- Ensure upstream and downstream integration for a synchronized supply chain quality network.

Some of the key aspects of a lean quality management system are:

1 – Risk-based approach emphasizing FDA quality priorities. For instance, for an ANDA filing emphasis should be on proving that the final product meets the quality of previous drug filings, not on proving the efficacy of the drug itself in clinical trials. The lean quality program will be digital and designed from the ground up to include: Process Validation Plan (PVP), User Requirement Specification (URS), Functional Requirement Specification (FRS), Test Protocols, Installation Qualification (IQ), Operational Qualification (OQ), and PQ (Performance Qualification (PQ). Quality program data and process data will be digitally captured, structured, and integrated to support the generation of a compliant digital regulatory framework aligned with FDA submission requirements."

2 – Enabling knowledge management and extracting information from a system. The quality and accessibility to the data makes it easier to understand the system entity relationships, to make it clear what is important, and to do all of this in realtime to make faster decisions to reduce risk.

3 – The digital model-based approach can help to reduce the effort required to maintain the system. Change control in qualified systems (under GMP) is often constrained by manual documentation and limited system visibility, especially in legacy environments. A digital system based on robust and validated models can lower the regulatory impact and validation requirements. If the data has been captured and organized well, changes to the system can be evaluated and approved faster and safer using digital analysis tools including AI. Changes made to the system can be more effectively managed at the plant level through improved traceability and data integrity and can be communicated to the FDA with greater clarity and supporting documentation, facilitating more informed and timely regulatory review.

In conclusion, Continua Process Systems has developed a digital platform that integrates a lean quality management system with a state-of-the-art AI enabled digital data integration

platform. This platform can significantly decrease the end-to-end time to design, build, qualify, and get FDA approval of new pharmaceutical plants for biologics, API, and OSD. This end-to-end plant deployment solution is an important part of the movement to onshore drug manufacturing to the US. Continua has a unique set of experiences and skills sets and can provide a turnkey solution for regulatory-ready, AI-powered pharmaceutical manufacturing.