Sarcouncil Journal of Applied Sciences

ISSN(Online): 2945-3437

Volume- 05| Issue- 09| 2025

Research Article **Received:** 30-07-2025 | **Accepted:** 20-08-2025 | **Published:** 16-09-2025

Automated Domain-Centric Data Products: Integrating Data Mesh Architecture with Robotic Process Automation in Pharmaceutical Enterprises

Sravish Nalam

BITS, Pilani, India

Abstract: Pharmaceutical companies are faced with growing data complexity in clinical, regulatory, and commercial areas, with old centralized architectures unable to provide the agility and scalability required for current drug development and commercialization. An emerging architecture utilizing Data Mesh principles with Robotic Process Automation (RPA) and DataOps provides a framework to build intelligent data product pipelines that are domain focused and allow domain teams to take ownership of their data products while also applying automated processing to repeatable activities such as data ingestion, metadata registration, quality validation, lineage verification, and compliance checks. The proposed architecture establishes self-service infrastructure capabilities and federated governance models that embed automation at critical integration points, facilitating continuous integration and deployment of trusted data products. Real-world implementations demonstrate significant improvements in data product delivery cycles, quality metrics, and regulatory compliance across clinical trial standardization, omnichannel analytics, and regulatory reporting workflows. The model or framework enables pharmaceutical organizations to meaningfully scale their data ecosystems with less manual overhead while still maintaining the important governance and quality controls critical for healthcare innovation. This architectural convergence signifies a shift from large monolithic data platforms to decentralized, more automated ecosystems that allow for quicker insight generation and position pharmaceutical organisations for large-scale AI-enabled innovation.

Keywords: Data mesh, robotic process automation, pharmaceutical data management, domain-centric architecture, DataOps.

INTRODUCTION

The Data Challenge in Pharmaceutical Enterprises

Current State of Pharma Data Management

The pharmaceutical sector is at a vital point at which data management capabilities will directly impact timelines to drug discovery, regulatory compliance, and commercial success. Today's pharmaceutical organizations are exhibiting new complexity in their data ecosystem from an ever-expanding and quickly changing collection of systems used in drug development, including clinical trials, regulatory affairs, manufacturing, and commercialization. These organizations usually support dozens of specialized databases, laboratory information management systems (LIMS), electronic data capture (EDC) systems, and enterprise resource planning (ERP) systems, each creating thousands of terabytes of structured and unstructured data. Accordingly, the data ecosystem is extremely fragmented with data ownership across functional silos and core business processes, relying primarily on humans to integrate, validate, and transform data.

The multiplicity of data sources has resulted in an environment where pharmaceutical companies have to reasonably expect that they are unable to maintain data consistency and quality in their operations. Most of the processes for data entry, validation, transformation, and reporting are manual; therefore, decision and reporting delays, errors, and compliance risks are inevitable. Domain teams often do not have direct access to their data sources, which eliminates ownership of the data, and rely on internal IT units that are bottlenecks in their data delivery pipeline. This fragmentation extends beyond technical systems to organizational structures, where unclear data ownership and accountability further complicate efforts to leverage data as a strategic asset (Imran, S. *et al.*, 2020).

Constraints of the Centralized Data Architecture

For many organizations, traditional centralized data architectures have been built upon enterprise data warehouses and data lakes. Traditional monolithic paradigms work well to solve common organization-wide business problems, but they fall short in dealing with the complexity of pharmaceutical data management. The centralized approach is encountering limitations when faced with the diverse types of data, specific regulatory constraints, and rapid iterations needed through modern analytics within pharmaceutical operations. The centralized data architecture approach creates bottlenecks and delays. It requires organizations to prioritize domain-specific innovation and compromises the diverse governance arrangements needed for clinical, regulatory, and commercial functions under the same structure.

Healthcare organizations that bring in big data and analytics face several hurdles when it comes to data integration, normalization, and governance,

***Corresponding Author:** Sravish Nalam

DOI- https://doi.org/10.5281/zenodo.17132469

especially when trying to adhere to a single source of truth approach across multiple domains (Imran, S. *et al.*, 2020). In the pharmaceutical context, these hurdles become even more pronounced by adding more complexity, such as subject to Good Clinical Practice (GCP) guideline adherence, requirements for regulatory submission, and the need for end-to-end traceability. Centralized architecture typically lacks the flexibility for real flexibility across multiple domains, so that all data consumers are forced to consume data through the same lens, and therefore, no one is fully happy with the one-size-fits-all approach. This creates a technical debt and organizational friction that hinders innovation and slows the pace of drug discovery and commercialization.

The Promise of Decentralized Approaches

The emergence of Data Mesh principles offers a compelling alternative to traditional centralized architectures, proposing a paradigm shift toward domain-oriented decentralization. This approach treats data as a product, assigns ownership to domain teams closest to the data source, and establishes federated governance while providing self-serve data infrastructure. The importance of data-driven enterprise architecture in pharmaceutical R&D becomes evident when considering how decentralized approaches can enhance agility and innovation while maintaining necessary controls (Uzhakova, N., & Fischer, S. 2024).

Through self-contained concerns, Data Mesh addresses the core of the tension between a centralized decision-making authority and dispersed innovation by empowering domain auto creation by using teams to independently grow, maintain, and enhance their data products, but at the same time, having to adhere to enterprise quality standards for interoperability. This is an important paradigm shift, recognizing that the domain expert has the best idea about the context of the data, the quality requirements, and the use cases. By putting ownership of the data in the hands of the domain expert and giving them the tools and platforms to manage their data products, organizations can drive innovation through the product life cycle, all the while maintaining the relevant governance and compliance protocols that are typically expected in the pharmaceutical world.

Research Objectives and Integrated Architecture Vision

We propose a unified architecture using Data Mesh, Robotic Process Automation (RPA), and DataOps for intelligent, domain-centered data product pipelines for pharmaceutical companies. The combined use of alluded technologies addresses significant gaps in current adoptions by automating repetitive data management activities, developing continuous integration and deployment workflows for data products, and providing scalable governance in decentralized domains.

While RPA technologies automate rules-based processes, such as data ingestion, quality validation, and compliance checks, DataOps practices help ensure the reliable, repeatable delivery of high-quality data products, as part of a data product ecosystem. This would support a complete reframing of how pharmaceutical organizations can build their data ecosystem to address their operational needs today and for future innovation opportunities. The architecture proposed within the document could reduce manual overhead, decrease time-to-insight, and lay the groundwork for AI-enabled drug discovery and development at scale.

Paper Structure and Contributions

The following sections provide a detailed framework to assist pharmaceutical companies in implementing this integrated architecture. Section 2 develops the building blocks by explaining how the Data Mesh, RPA, and DataOps paradigms intersect and how they can be combined through an integrated architecture. Section 3 describes the architecture, including self-service components and federated governance models to create the right balance between autonomy and controls. Section 4 demonstrates the automation use cases and implementation approaches for a pharmaceutical context, enabling organizations to leverage the guidance in this document for their transformation. Section 5 provides case studies in clinical data standardization, commercial analytics, and regulatory reporting to showcase how the proposed architecture has been successfully put in place. The article concludes with commentary on the future of automated data ecosystems for autonomous pharmaceutical innovation and the speed-to-value advantage such ecosystems can provide for drug development and commercialization, and the zero-risk barrier regarding data quality and regulatory compliance.

Publisher: SARC Publisher

Table 1: Comparison of Traditional vs. Data Mesh Architecture in Pharma (Imran, S. *et al.*, 2020; Uzhakova, N., & Fischer, S. 2024)

Aspect	Traditional Centralized Architecture	Data Mesh Architecture
Data Ownership	Centralized IT/Data Team	Domain Teams
Scalability	Vertical (Limited)	Horizontal (Unlimited)
Time to Insight	Weeks to Months	Hours to Days
Domain Expertise Integration	Disconnected from Data Management	Embedded in Data Product Development
Governance Model	Centralized Control	Federated with Computational Policies
Innovation Speed	Slow due to Bottlenecks	Rapid through Domain Autonomy
Compliance Management	Manual Reviews and Audits	Automated Continuous Validation

THEORETICAL FOUNDATION: DATA MESH, RPA, AND DATAOPS CONVERGENCE

Data Mesh Principles

Data Mesh exemplifies a shift in data architecture from centralized data platforms to distributed data ownership in the form of domain-oriented data architectures. The conceptual framework for data mesh is based on four principles, each altering the way that organizations think about the management and utilization of their data assets. Domain ownership is an important proponent that imparts ownership and accountability to the party that has the most context of the data and its narratives, which modifies the traditional concept of clearly delineated teams with centralized data responsibility for all enterprise data. The domain ownership principle realizes that an organization's domain experts will always have a context of their data's quality, narratives, and nuances that no central organization can obtain. (Butte, V. K., & Butte, S. 2022).

The concept of data as a product elevates data from a byproduct of operational systems to a first-class citizen with defined consumers, quality standards, and service level agreements. Domain teams apply product thinking to their data, considering user experience, documentation, discoverability, and reliability as core attributes. Self-serve data infrastructure provides domain teams with the platforms and tools needed to create, maintain, and evolve their data products independently, eliminating dependencies on centralized teams for routine operations. Federated governance balances the need for enterprise-wide standards and compliance with domain autonomy, establishing computational policies that enforce standards while allowing domains to innovate within defined boundaries (Butte, V. K., & Butte, S. 2022)

RPA Capabilities in Data Pipeline Automation

Robotic Process Automation is a strong enabler of automating repetitive, rules-based tasks within data pipelines that can provide pharmaceutical organizations with tremendous opportunities to optimize their data activities. RPA technologies are particularly effective in that they can replicate human interactions with digital systems and automate users' lighter, repetitive workflows consistently and reliably, which manual processes cannot match. In the context of data pipelines, RPA bots can traverse multiple systems, extract data from multiple sources, perform validation checks, and apply transformation rules autonomously (Czarnecki, C., & Fettke, P. 2021).

RPA has great potential for integration throughout the data lifecycle from data capture to delivery into the systems that utilize the data. RPA platforms have capabilities for screen scraping, API interactions, file actions, and database actions, which allow for integration with legacy systems that have outdated APIs or none at all. These capabilities are especially useful in the pharmaceutical environment, where data systems can be decades old, with new systems and manual data movement from one system to another still being common. The rule-based approach of RPA is consistent with the structured, regulated processes that are typical of pharmaceutical data management, where compliance dictates consistent execution of processes that are defined (Czarnecki, C., & Fettke, P. 2021).

DataOps Methodology

DataOps has adapted the principles of agile and DevOps to data analytics to create disciplines that integrate to provide continuous delivery of operational, high-quality data products. DataOps methodology (Deswandikar, A. 2024) puts a strong emphasis on automation, collaboration, and continuous improvement across the data lifecycle, and encourages the use of CI/CD pipelines for data

Publisher: SARC Publisher

products. CI/CD pipelines enable teams to make changes quickly while ensuring quality through automated testing and validation. CI/CD pipelines generally involve data ingestion, data transformation, data quality checks, and data production deployment to help safeguard safe, predictable changes to the data product.

Within DataOps, "quality assurance" for data products goes beyond data validation to take into account statistical process control (SPC), anomaly detection, and automated reconciliation. Monitoring not only enables a rich source of information on data pipeline performance, data quality metrics, and system health, and includes capabilities to swiftly find and fix issues. DataOps enables individual data engineers and teams, domain experts, and end data consumer groups to collaborate on models and tailored outputs with access to common tools and transparent processes, through defined roles and channels of communication, thus enabling safety and transparency, while maintaining regulatory standards.

Synergies Among the Three Paradigms

The relationship between Data Mesh, RPA, and DataOps creates synergies that can address the challenges associated with pharmaceutical data management. Data Mesh provides the organizational and architectural approach to provide distributed ownership of data, and RPA removes some of the repetitive activities that otherwise strain domain teams. DataOps ensures that the data products made available will be trustworthy and of good quality by implementing processes that are automated when possible. Collectively, these paradigms allow pharmaceutical organizations to build their data capabilities at scale and still fulfil their governance and compliance responsibilities.

RPA helps enable Data Mesh by automating the operational responsibilities when managing data products. Domain teams can use RPA to automate ingestion from data source systems, perform routine assessments of data quality, manifest metadata for data cataloging, track data lineage, and catalogue data-provenance without any human intervention. DataOps provides institutional knowledge to ensure these automated processes become established, robust processes, so that they will safely evolve through version control, testing, and deployment pipelines. Together, they allow domain teams to focus on better understanding the data products they develop and improving them rather than monitoring operating tasks.

Table 2: Data Mesh Principles and RPA Integration Points (Butte, V. K., & Butte, S. 2022; Czarnecki, C., & Fettke, P. 2021)

Data Mesh Principle	RPA Automation Capability	Pharmaceutical Application
Domain Ownership	Automated operational tasks	Clinical teams manage trial data products
Data as a Product	Quality validation and testing	SDTM datasets with automated QC
Self-Serve Platform	Infrastructure provisioning	Domain teams deploy data pipelines
Federated Governance	Policy enforcement and monitoring	GxP compliance validation

Related Work and Existing Approaches

The pharmaceutical industry has experimented with a range of approaches to improve data management, from the foundational principles of data warehousing to more up-to-date data lake approaches. As outlined in earlier research, the initial focus was on combining clinical trial data, manufacturing data, and commercial operations data into a single repository. This approach was relatively successful at providing limited available views of data to help breakdown silos, yet struggled with scalability and agility. More recent approaches have begun to explore cloud-based platforms and advanced analytics, but organizations continue to find it troublesome integrating multiple raw data sources, establishing data quality and consistency, and considering data management in terms of regulatory requirements.

Existing literature shows consistent interest in leveraging modern data architecture patterns in pharmaceutical content, but fewer examples demonstrate fully implementing the Data Mesh principles combined with automation technologies. Organizations have explored data systems with a domain-driven design, or have explored automation with RPA limited to certain use-cases (e.g., regulatory reporting, clinical data processing). However, the simultaneous integration of all three fields (Data Mesh, RPA, and DataOps) is uniquely ambiguous, in that it's a new way of thinking about the limitations of past implementations. The convergence of all three presents a pathway that can combine organizational change and technological change to create sustainable and scalable data ecosystems

Publisher: SARC Publisher

suited to support the complex needs of pharmaceutical enterprises.

PROPOSED ARCHITECTURE

High-Level Architectural Blueprint

The proposed architecture integrates Data Mesh principles with Robotic Process Automation to create an intelligent, scalable framework for pharmaceutical data management. At its core, the architecture establishes autonomous domain teams as primary data product owners while providing them with automated tools and platforms to manage their responsibilities effectively. The blueprint envisions a federated ecosystem where clinical, regulatory, manufacturing, and commercial domains operate independently yet interoperably, connected through standardized interfaces and shared infrastructure components (Deswandikar, A. 2024).

The architectural design positions RPA as an integral layer within the Data Mesh framework, automating routine operations that would otherwise consume significant domain team resources. This integration occurs at multiple levels, from infrastructure provisioning to data pipeline execution, creating a self-sustaining ecosystem where automation enables rather than replaces human expertise. Domain teams retain control over their data products' logic and business rules while RPA handles execution, monitoring, and operational tasks. The architecture leverages cloud-native technologies to provide elasticity and scalability, ensuring that domains can grow their data capabilities without architectural constraints (Deswandikar, A. 2024).

Self-Service Infrastructure Components

Self-service infrastructure forms the foundation of the proposed architecture, empowering domain teams with the tools and platforms necessary to create, manage, and evolve their data products independently. The infrastructure layer provides standardized components, including data storage services, compute resources, orchestration engines, and development environments, all accessible through intuitive interfaces that abstract underlying complexity. Domain teams can provision resources, deploy data pipelines, and manage their data products without requiring deep technical expertise in infrastructure management.

Platform capabilities extend beyond basic infrastructure to include pre-built templates for common data product patterns, automated testing frameworks, and integrated development environments tailored for data work. These capabilities accelerate data product development while ensuring consistency across domains. The platform provides APIs and SDKs that enable domains to extend functionality while maintaining compatibility with enterprise standards. Integration with existing pharmaceutical systems occurs through standardized connectors and adapters, allowing domains to access source systems without managing complex integration logic (Mahendran, P. K. R. 2024).

Federated Governance Model with Embedded Automation

The federated governance model balances domain autonomy with enterprise-wide standards through computational policies and automated enforcement mechanisms. Governance policies encode regulatory requirements, data quality standards, and security protocols as executable rules that RPA bots enforce consistently across all domains. This approach transforms governance from a manual review process to an automated, continuous validation system that operates in real-time as data products evolve.

Embedded automation within the governance framework includes automated policy validation, compliance checking, and audit trail generation. RPA bots monitor data product changes, validate them against governance policies, and either approve changes automatically or escalate exceptions to human reviewers. The model supports different governance tiers, allowing critical data products to receive more stringent oversight while enabling rapid innovation for experimental or low-risk products. Governance automation extends to documentation requirements, with RPA generating and maintaining compliance artifacts automatically based on data product metadata and lineage information (Mahendran, P. K. R. 2024)

RPA Integration Points

RPA integration occurs at critical junctures throughout the data product lifecycle, automating tasks that traditionally require manual intervention. Data ingestion automation handles the extraction of data from source systems, whether through APIs, file transfers, or screen scraping from legacy applications. RPA bots execute ingestion schedules, monitor for data availability, handle exceptions, and validate incoming data against expected schemas. This automation ensures consistent, reliable data flow into domain data products without constant human oversight.

Publisher: SARC Publisher

Metadata management represents another crucial integration point where RPA automates the capture, validation, and maintenance of data product metadata. Bots automatically extract technical metadata from data sources, enrich it with business context from documentation systems, and maintain the enterprise data catalog. Quality validation automation executes comprehensive data quality checks, comparing data against business rules, statistical thresholds, and historical patterns. Lineage tracking automation captures data transformation logic, system interactions, and processing steps, maintaining complete audit trails for regulatory compliance. These integration points work in concert to create end-to-end automation of data product operations.

DataOps Implementation

DataOps implementation within the architecture establishes continuous integration, testing, and deployment workflows specifically designed for data products. The CI/CD pipeline begins with version control for all data product artifacts, including transformation code, configuration files, and infrastructure definitions. Automated testing encompasses unit tests for transformation logic, integration tests for data pipeline components, and regression tests to ensure changes don't impact downstream consumers.

Deployment workflows apply their infrastructure as code principles and allow the domains to define their data product infrastructure declaratively and make changes through automated and efficient pipelines. There's built-in support for automated rollback, blue-green deployments for zero-downtime updates, and canary releases for controlled releases. Monitoring and observability tools give real-time insight into a variety of metrics, including, but not limited to, pipeline performance, data quality metrics, and health metrics for systems, which enables proactive identification and remediation of issues. The DataOps framework provides stability and predictability so that data products can evolve safely while still being highly available and reliable.

Security and Compliance Considerations

Security and compliance considerations touch every layer of the proposed architecture; the good security risk management of information is no surprise given the strict requirements for managing pharmaceutical data. The architecture uses defense-in-depth security strategies, including network segmentation, encryption both at rest and in transit, and takes a fine-grained approach towards access rights management. RPA bots work with organizational rights and privileges, the RPA bots only access systems and data that are necessary to perform limited tasks. Various authentication and authorization mechanisms integrate with enterprise identity providers, providing consistent access across domains.

Compliance automation manages regulatory requirements, including GCP, and data privacy laws like GDPR and HIPAA. RPA Bots automatically enforce data retention policies, manage consent tracking, and do data anonymization when it is required. The complete architecture records all the access and access transformations for the data within an Allan Tanner audit log. These provide the traceability that regulatory officials consistently ask for review during investigations. Compliance automation continuously verifies compliance, alerting and spinning off remediation workflows immediately when exceptions occur. The security and compliance posture of the proposed system and architecture solution keeps the chaotic decentralized nature of Data Mesh while preserving control and auditing requirements that apply to any pharmaceutical environment.

AUTOMATION PATTERNS AND IMPLEMENTATION FRAMEWORK

Core RPA Automation Patterns for Pharma Data Products

The implementation of RPA within pharmaceutical data products follows distinct patterns that address industry-specific requirements while maximizing automation benefits. Event-driven automation patterns trigger RPA workflows based on system events such as file arrivals, database updates, or scheduled intervals, ensuring the timely processing of critical data. These patterns incorporate exception handling mechanisms that gracefully manage errors while maintaining data integrity and compliance requirements. Orchestration patterns coordinate multiple RPA bots working in sequence or parallel, enabling complex workflows that span multiple systems and data domains (Yadav, N., & Panda, S. P. 2022).

Template-based automation patterns standardize common pharmaceutical data processes such as adverse event reporting, clinical trial data collection, and regulatory submission preparation. These templates encapsulate best practices and

Publisher: SARC Publisher

regulatory requirements, allowing domain teams to implement compliant automation rapidly. The patterns include configurable parameters that adapt to specific domain needs while maintaining consistency across the enterprise. Adaptive automation patterns leverage machine learning capabilities to improve process efficiency over time, learning from exceptions and adjusting processing rules within defined boundaries to handle variations in data formats and content (Yadav, N., & Panda, S. P. 2022)

Table 3: RPA Automation Patterns for Pharmaceutical Data Products (Yadav, N., & Panda, S. P. 2022)

Pattern Type	Description	Pharmaceutical Use Cases
Event-Driven	Triggered by system events	Lab result processing, AE reporting
Template-Based	Standardized process templates	SDTM mapping, regulatory submissions
Orchestration	Multi-robot coordination	End-to-end clinical data flow
Adaptive	ML-enhanced processing	Data format variations, exception handling
Scheduled	Time-based execution	Daily safety reports, periodic reconciliation

Metadata Registration and Catalog Automation

Automated metadata registration transforms the traditionally manual process of documenting data products into a continuous, accurate system of record. RPA bots automatically extract technical metadata from data sources, including schema definitions, data types, constraints, and relationships. The automation extends to capturing business metadata through integration with documentation systems, extracting definitions, ownership information, and usage guidelines from various repositories. This comprehensive metadata collection ensures that the enterprise data catalog remains current and complete without requiring manual updates.

The catalog automation framework implements intelligent classification algorithms that categorize data products based on content analysis, usage patterns, and regulatory requirements. RPA bots enrich metadata with additional context, including data quality scores, usage statistics, and compliance status, providing consumers with comprehensive information for decision-making. The automation includes version control for metadata changes, maintaining historical records of how data products evolve. Integration with data discovery tools enables automatic tagging and indexing, improving searchability and facilitating data product reuse across domains.

Data Quality Check Automation and Anomaly Detection

Data quality automation implements multi-layered validation strategies that ensure pharmaceutical data meets stringent accuracy and completeness requirements. RPA bots execute syntactic validations, checking data format, structure, and type conformity against defined schemas. Semantic validations verify that data values fall within acceptable ranges and maintain referential integrity across related datasets. Statistical quality checks monitor data distributions, identifying outliers and anomalies that may indicate quality issues or processing errors.

Anomaly detection patterns leverage both rule-based and statistical methods to identify data quality issues proactively. Time-series analysis identifies trends and seasonality in data patterns, flagging deviations that require investigation. Cross-domain validation ensures consistency when data appears in multiple systems, automatically reconciling discrepancies and generating exception reports. The framework implements feedback loops where confirmed anomalies update detection rules, continuously improving the system's ability to identify quality issues. Real-time quality dashboards provide domain teams with immediate visibility into data quality metrics, enabling rapid response to emerging issues.

Compliance Validation Workflows

Compliance validation workflows can be set up to fully automate the process of determining whether data products are compliant with regulatory criteria (e.g., Good Clinical Practice (GCP), Good Manufacturing Practice (GMP), Good Laboratory Practice (GLP), etc.). The RPA bots can execute validation processes that include all necessary validation procedures encompassing completeness, accuracy, and traceability as part of a given product's full data lifecycle. Compliance validation workflows include regulatory-specific checks such as the identification of protocol deviations in GCP, completeness checks for batch records in GMP, and chain of custody checks for laboratory data (Samson, Y. 2006).

Privacy compliance automation addresses HIPAA, GDPR, and other data protection regulations through systematic validation of data handling practices. RPA bots verify that personal health

Publisher: SARC Publisher

information undergoes appropriate de-identification, that consent management processes operate correctly, and that data retention policies execute as configured. The workflows generate compliance certificates and exception reports automatically, maintaining audit-ready documentation for regulatory inspections. Automated compliance monitoring operates continuously, flagging potential violations before they escalate into regulatory issues. The framework adapts to regulatory changes through configurable rule sets that update without requiring extensive system modifications (Samson, Y. 2006).

Lineage Tracking and Audit Trail Automation

Comprehensive lineage tracking automation captures the complete journey of data from source systems through all transformations to final consumption points. RPA bots automatically document data sources, transformation logic, quality checks performed, and systems accessed during processing. This automation creates detailed audit trails that satisfy regulatory requirements for data traceability while providing operational insights into data flow patterns. The lineage system captures both technical transformations and business process contexts, linking data changes to specific business events or decisions.

Audit trail automation goes beyond transforming data and encompasses all activities within the system by the user and automated processes. RPA bots pull audit information from various systems, allowing a complete audit trail in the data ecosystem. This automated audit trail includes intelligent compression and archiving strategies that allow for full audit histories of any data while resolving storage issues. Audit monitoring can identify anomalous behavior that may indicate suspicious activity or unauthorized access from users or automated processes, which can trigger alerts and release a response to address concerns. This framework enables an unalterable audit trail through the critical infrastructure of cryptographic signatures and immutable storage.

Platform APIs and Integration Standards

Standardized APIs form the core for the automation framework and facilitate component, data platform, and domain system interoperability. RESTful APIs enable synchronous interfaces to access, in real-time, control operations, and event-driven APIs enable asynchronous processing patterns. The API Framework implements authentication, authorization, and rate limiting to afford secure and managed access to automation services. Versioning strategies ensure backward compatibility while enabling API longevity and preventing existing integrations from being disrupted.

Integration standards define how to share data, consume data from, respond to errors and events, and monitor integration in the automation ecosystem. Integration standards involve message format specifications, protocols for interfacing, and service-level agreements that ensure reliable communication between systems. Integration patterns can include point-to-point connections for simple integration or enterprise service bus architecture for complex multi-system workflows. Adapter libraries provide pre-packaged connectors to common pharmaceutical systems, allowing for software accelerants to be designed while maintaining consistency in the integration process. Integration standards will also include ways to integrate with legacy systems, using screen scraping, and file-based interfaces where APIs do not exist.

Domain Team Enablement and Change Management Strategies

Comprehensive enablement programs are necessary for successful implementation that will empower domain teams to take advantage of automation capabilities. Training activities involve hands-on workshops and self-service learning opportunities, allowing teams to acquire automation skills over time. The enablement framework includes an automation champion in each domain to act as local expertise and change agents to convey knowledge and support adoption. Regular showcases of successful automation implementations provide teams with the inspiration to recognize new opportunities for automation in their domains.

Change management strategies encompass both the technical and organizational changes that are necessary for the successful adoption of automation. Phased rollout strategies enable phases to adopt automation incrementally as they build confidence in the solution from early successes before expanding the scope. The strategy includes metrics and KPIs that quantify the value of automation, increasing the likelihood of organizational support for future investment. Communications plans ensure that there is transparency through the implementation process about the possible implications of job loss as

Publisher: SARC Publisher

existing positions are augmented rather than replaced by humans when achieving navigator status; their human reasoning and input will now be much more valued. Governance structures establish distinct roles and responsibilities for the development, maintenance, and evolution of automation, allowing for continued long-term sustainable success.

REAL-WORLD APPLICATIONS AND CASE STUDIES

Clinical Trial Data Standardization: Automating SDTM Transformation Pipelines

The implementation of automated Study Data Tabulation Model (SDTM) transformation pipelines represents a critical advancement in clinical trial data management. Traditional SDTM mapping processes require extensive manual effort from clinical data managers who must interpret protocol requirements, map source data to SDTM domains, and validate the resulting datasets. The automated approach leverages RPA bots that execute predefined mapping rules, transform raw clinical data into SDTM-compliant formats, and perform comprehensive validation checks against CDISC standards. This automation significantly reduces the time required for database lock while improving consistency across studies (Kim, S. 2024).

The architecture implements intelligent mapping templates that adapt to different therapeutic areas and study designs while maintaining SDTM compliance. RPA bots automatically identify source data elements, apply transformation logic, and generate annotated CRFs that document the mapping process. The system incorporates machine learning algorithms that learn from previous mapping decisions, suggesting optimizations and identifying potential issues before they impact submission timelines. Quality control automation validates transformed data against SDTM implementation guides, generating detailed compliance reports and identifying discrepancies that require manual review. The framework supports incremental data processing, enabling continuous SDTM generation throughout the study lifecycle rather than waiting for a database lock (Kim, S. 2024).

OMNICHANNEL COMMERCIAL ANALYTICS

Unifying Marketing and Sales Data Products

Pharmaceutical commercial teams increasingly require integrated views across multiple customer engagement channels, including field sales, digital marketing, medical affairs, and patient support programs. The implementation of unified commercial data products addresses this need by automating the integration of disparate data sources into cohesive analytical assets. RPA bots extract data from CRM systems, marketing automation platforms, web analytics tools, and call center databases, harmonizing different data formats and schemas into standardized data products that support omnichannel analytics.

The automation framework implements sophisticated entity resolution algorithms that match healthcare providers, institutions, and patients across different systems while maintaining privacy compliance. Domain teams in marketing, sales, and medical affairs maintain ownership of their respective data products while contributing to shared commercial intelligence assets. The architecture enables real-time synchronization of customer interactions, ensuring that all teams work with current information when engaging healthcare stakeholders. Advanced analytics built on these unified data products reveal cross-channel attribution patterns, optimize resource allocation, and identify opportunities for personalized engagement strategies. The system maintains compliance with promotional guidelines by automatically tracking and validating all customer interactions against approved messaging.

Regulatory Reporting Automation: Streamlining Submission-Ready Datasets

Regulatory submission preparation traditionally involves labor-intensive processes to compile, validate, and format datasets according to health authority requirements. The automated approach transforms this process through RPA-driven workflows that generate submission-ready packages with minimal manual intervention. The system integrates with clinical data management systems, safety databases, and document repositories to automatically compile required datasets, generate define.xml files, and create reviewer guides. Automation extends to the validation of submission packages against regulatory technical specifications, ensuring completeness and compliance before submission (Kamaraj, A., & Puranik, S. 2023).

The framework adapts to different regulatory requirements across global health authorities, maintaining region-specific formatting and validation rules. RPA bots execute comprehensive cross-checks between submitted data and source

Publisher: SARC Publisher

systems, identifying discrepancies that could delay regulatory review. The automation includes intelligent scheduling capabilities that coordinate data extraction and compilation activities based on submission timelines, ensuring that teams meet regulatory deadlines without last-minute scrambles. Post-submission automation monitors health authority portals for feedback and automatically routes queries to appropriate domain experts for response. The system maintains complete audit trails of all submission preparation activities, facilitating regulatory inspections and demonstrating data integrity (Kamaraj, A., & Puranik, S. 2023).

Performance Metrics and KPIs

Automating the data product pipeline produces observable improvements along many dimensions of performance. Cycle time metrics have significantly decreased the durations of time to produce analytical datasets; automation overcomes the delays associated with manual processing and provides near-real-time access to data. Data quality metrics show marked improvements in accuracy, completeness, and consistency as automated data validation automates many errors that are missed through a manual process. The framework measures operational metrics such as bot handling, exception handling efficiency, and uptime to verify that the operational system is functioning properly.

Business value metrics calculate the effect of automation on some aspects of pharmaceutical operations, such as start-up times from clinical trials, faster regulatory submissions, and quicker commercial decisions. The cost savings measurements are based on reduced manual effort, reduced costs for dealing with errors, and avoiding regulatory delays. The system uses automated dashboards that provide real-time visibility to those metrics and help us improve continuously through optimizing automation. A predictive analytics capability using historical performance data can help identify bottlenecks and suggest improvements to the process. This creates a cycle of continuous improvement and maximises automation value.

Lessons Learned and Best Practices

Implementation experiences across several pharmaceutical organizations suggest several critical success factors for automation. With those success factors in mind, organizations should target initial uses of automation to areas with easy-to-define, high-volume processes that allow for early wins that can build comfort with how and why to automate. Introducing governance structures and expectations as part of the planning and preparation ensures confusion surrounding expectations and roles is minimized when automations require a larger implementation across domains. Also, investing in training programs that inform users of the range of possibilities with automation, as well as the limitations of automation, will be a central component to sustainable adoption, as teams that own the solution will need to understand what automation technologies can offer.

Technical lessons emphasize the importance of modular architecture that allows incremental adoption and easy modification as requirements evolve. Standardizing data formats and interfaces early in the implementation process significantly reduces integration complexity as new domains join the ecosystem. Regular assessment of automation performance against defined success criteria ensures that implementations deliver expected value. Organizations that maintain flexibility in their automation strategies adapt more successfully to changing regulatory requirements and business needs. The most successful implementations balance automation sophistication with maintainability, avoiding overly complex solutions that become difficult to sustain.

Scalability Considerations and AI-Readiness Assessment

Scalability planning addresses both technical and organizational dimensions of growing automation capabilities across the enterprise. Technical scalability requires cloud-native architectures that elastically expand to handle increasing data volumes and processing demands. The framework implements horizontal scaling patterns that distribute processing across multiple RPA bots and computing resources, preventing bottlenecks as automation scope expands. Organizational scalability depends on federated operating models that empower domains to develop automation capabilities independently while maintaining enterprise coherence through shared standards and platforms.

An AI-readiness assessment focuses on the essential layer that automated data products offer, upon which all advanced analytics and machine learning initiatives take place. High-quality and automated data products that are standardized create the perfect datasets to train the AI models. Automated capture of all metadata and lineage also

Publisher: SARC Publisher

allows the AI system to have context and develop accurate predictions. The integration pattern, which is pre-established for the RPA, also generates automated pathways to embed and utilize AI in real-world operational workflows, resulting in intelligent automation that merges rules-based processing with machine learning prediction capabilities. Organizations that have successfully implemented these layers of the proposed architecture for the use of analytics will be best positioned to harness attention and funding for new AI technologies for drug discovery, optimizing clinical development, and personalized medicine.

CONCLUSION

The intersection of the principles of Data Mesh, Robotic Process Automation, and DataOps constitutes a fundamental change in the language of a pharmaceutical's data architecture and management. In particular, we move from centralized architectures to intelligent ecosystems focused on domains. By evolving this architecture to address core challenges faced by pharmaceutical companies, domain teams are empowered with automated capabilities while ensuring compliance and governance policies essential to drug development and commercialization. The case studies and implementation stories that we share demonstrate that when companies fully embrace this integrated data architecture framework, they are able to significantly improve the speed, quality, and reliability of data products while lowering operational overhead and human error related to manual data processing. Moreover, the automated data product pipelines provided pharmaceutical companies with the ability to leverage their data much more intelligently and expediently, thus shortening clinical timelines, easing regulatory burden, and becoming agile with commercial decisions. As the pharmaceutical sector continues its journey of digital transformation, we believe that the advances made through establishing automated and decentralized data architectures have enabled organizations to the point of implementing and scaling for new and emerging technologies such as artificial intelligence and machine learning. Future directions will involve increasing complex automation, combining rules-based processing with intelligence algorithms to create self-optimizing data ecosystems that adapt to changes in business need, all while assuring regulatory compliance. Organizations investing in developing these capabilities today will be better positioned to handle the complexities of modern drug development, deliver innovative therapies to patients even faster, and maintain a competitive advantage in an increasingly data-driven, high-tech healthcare market. Successfully deploying and integrating these technologies should not only be viewed as a technical achievement but as a complete transformation of how pharmaceutical companies can organize, manage, and gain value from their data assets in pursuit of improving patient outcomes.

REFERENCES

1. Imran, S., Mahmood, T., Morshed, A., & Sellis, T. "Big data analytics in healthcare− A systematic literature review and roadmap for practical implementation." *IEEE/CAA Journal of Automatica Sinica* 8.1 (2020): 1-22.
2. Uzhakova, N., & Fischer, S. "Data-driven enterprise architecture for pharmaceutical R&D." *Digital* 4.2 (2024): 333-371.
3. Butte, V. K., & Butte, S. "Enterprise data strategy: a decentralized data mesh approach." *2022 International Conference on Data Analytics for Business and Industry (ICDABI)*. IEEE, (2022).
4. Czarnecki, C., & Fettke, P. "Robotic Process Automation: Management, Technology, Applications." *IEEE Xplore*, (2021). https://ieeexplore.ieee.org/book/10783691
5. Deswandikar, A. “Engineering Data Mesh in Azure Cloud: Implement Data Mesh Using Microsoft Azure's Cloud Adoption Framework”. *Packt Publishing Ltd*, (2024).
6. Mahendran, P. K. R."Robotics–Integrating RPA with AI–Enhancing Decision-Making Process." *International Journal of Engineering Research & Technology (IJERT)*, August (2024). https://www.ijert.org/research/robotics-integrating-rpa-with-ai-enhancing-decision-making-process-IJERTV13IS080039.pdf
7. Yadav, N., & Panda, S. P. "A path forward for automation in robotic process automation projects: Potential process selection strategies." *2022 International Conference on Machine Learning, Big Data, Cloud and Parallel Computing (COM-IT-CON)*. Vol. 1. IEEE, (2022).
8. Samson, Y. "Computerized System Validation: Regulatory Compliance and Process Safety in the Pharmaceutical Industry." *Conference ERTS'06*. (2006).
9. Kim, S. "Digitalization and Data Standardization in Clinical Trials." *Korean Society for Clinical Development (KSCD)*,

Publisher: SARC Publisher

November (2024). https://www.cdisc.org/sites/default/files/2024-11/2024_cdisc-keynote_sanghee_kim_novotech_final.pdf

10. Kamaraj, A., & Puranik, S. "Automation in pharmaceutical industry and global regulatory compliance." *Journal of Current Pharma Research* 17.4 (2023): 13-25.

Source of support: Nil; **Conflict of interest:** Nil.

Cite this article as:
Nalam, S. " Automated Domain-Centric Data Products: Integrating Data Mesh Architecture with Robotic Process Automation in Pharmaceutical Enterprises." *Sarcouncil Journal of Applied Sciences* 5.9 (2025): pp 20-31.

Publisher: SARC Publisher