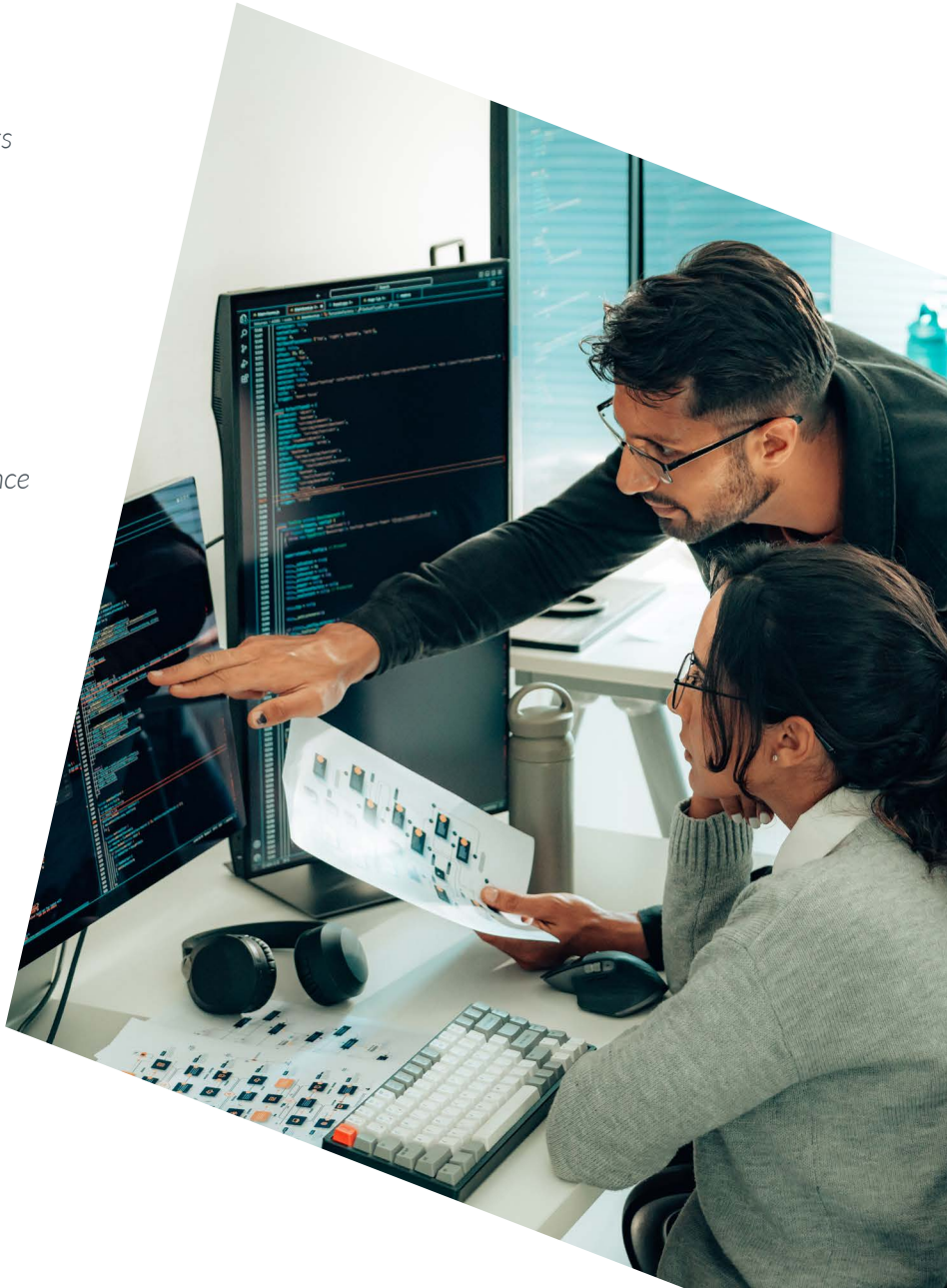
Delivering AI-managed clinical research data

Parexel's end-to-end automation strategy

The transformative technologies of artificial intelligence (AI) promise both greater management efficiencies and deeper insights into clinical trial data. Parexel envisions an AI-enabled platform that automates data management processes end to end—from data input, through analysis, to regulatory submission. Such a high-functioning, AI-powered data ecosystem must be underpinned by data governance principles and a technology framework that can advance data quality, acquisition and integration. This paper discusses Parexel's AI-enabled data strategy, including the development of its custom SDTM transformation tool in partnership with Palantir.



AI in data standardization: An ambitious vision

The drug development community has high expectations for the adoption of artificial intelligence (AI) to dramatically improve clinical trial operations, advance trial design, accelerate timelines and deliver better research outcomes. Foundational to this progress is AI-enabled automation of clinical data management, analysis, and reporting.

Sponsors are beginning to test the feasibility of machine learning technologies, including generative AI and natural language processing, to streamline the time- and labor-intensive tasks needed to manage increasing volumes and types of clinical trial data. Early results from first-generation AI solutions show improvements in speed, accuracy and utility for informed decision-making.[1] The vision for AI in drug development is to deploy AI-enabled processes that will deliver these advances across the full research and development value chain.

At Parexel, delivering on this promise requires a robust technology foundation that integrates all clinical data—from collection through analysis—in a connected, clinical data ecosystem. With processes automated by AI and overseen by human experts, this ecosystem will drive operational efficiencies and reduce development time. Real-time access and interoperability will leverage standardized data for earlier, more informed decision-making at critical points across the development pipeline.

Under construction: Parexel's AI-enabled ecosystem

To design this end-to-end AI-managed ecosystem, Parexel is using a data governance strategy that aligns people, processes and digital architecture. Alignment is guided by four tenets: framework, standards, tools, and people.

Framework. The framework must have the capability to manage *new types of data*—real-world data including electronic health records (EHRs), patient and healthcare professional texts and voice communications. It must be able to handle *larger volumes of data* as clinical evaluation expands to address drug impacts in real-world medical use. The framework must provide *real-time visibility* into study operations—a critical capability for effective stakeholder collaboration and streamlined operations. Key components include digital security; a cloud-centric data lake; and data streaming and sharing technologies. Comprehensive data governance will define and direct: data collection strategies; data standards; data integration methods; and the overall mapping of data distribution, security, protection and alignment with regulations.

Standards. Consistent data standards are necessary to drive operations and data quality across siloed and complex research processes. Standards make it possible to connect data and systems for decision making and streamlined execution of operations ranging from regulatory functions to reimbursement strategies and ongoing safety stewardship. A standards-driven approach enhances data value by improving data storability, aggregation and retrieval—both for immediate application and for archival uses.

Tools. Clinical development operations have evolved point-to-point tools that preserve siloes and pose barriers to interoperability and collaboration. Advances depend on developing a *source-agnostic*, AI-enabled platform that manages data using point-to-hub tools that reduce duplication and ensure a single source of truth.

People. AI-driven advances hinge on workforce readiness. The heavy responsibilities for quality and patient safety in therapeutic innovation demand *human-in-the-loop applications* that combine the speed and accuracy of automation with human expertise and oversight. The implementation of AI-human collaborations in everyday operations will require technical training and assimilation of changing roles and responsibilities. Parexel has a highly adaptive culture experienced in quickly operationalizing the rapid changes spurred by evolving AI adoption.

Executing the strategy: Partnership with Palantir

To build an end-to-end AI ecosystem, Parexel is partnering with Palantir Technologies Inc, a software company that builds platforms to power real-time, AI-driven decision-making. Palantir's core product, *AIP*, is used to create a central operating system that combines AI automation with human-in-the-loop decision making.[2]

AIP integrates high volume, multi-modal data into the Ontology, an AI-accessible environment that combines data and logic to enable rapid development of operational applications. AIP's robust support for standards management orchestrates rigorous evaluations of AI-produced content and guides human-in-the-loop interactions, enabling the application of AI to highly regulated workflows. It has capabilities to coordinate the provisioning of very sensitive data—with rights- and role-based access and security governance—to help ensure clinical trial data won't be unblinded accidentally or shared with unauthorized parties. Parexel began using AIP to integrate large volumes of unstructured real-world data, and quickly realized that value extended far beyond data integration—it provides a tool to rapidly build AI applications that leverage data to automate functions and streamline workflows.

Today, Parexel is using AIP as the engine to transform trial execution from silos of semi-manual tasks to highly efficient AI-powered workflows. SDTM transformation is the backbone of this revolution.

SDTM transformation: A ripe opportunity for AI

SDTM (Study Data Tabulation Model) is a set of standards developed by the Clinical Data Interchange Standards Consortium (CDISC). SDTM dictates the organization and formatting of clinical trial data to enable data sharing in the clinical research space. SDTM ensures accuracy, quality, and consistency—across studies and multiple stakeholders—and is a mandatory requirement for regulatory submissions globally.

Data generated in clinical trials must be converted (transformed) to SDTM format; this initial task in data management feeds downstream operations culminating in statistical analysis and regulatory submission. SDTM transformation is labor intensive and time consuming. Traditional methods rely on an array of custom programs to extract and manually map data to SDTM domains. It relies on manual data transfer, which is both costly and prone to error; extensive quality checks are necessary to reconcile changes in source data, flag missing data points, and address other anomalies.

Growing volumes of clinical data and types of inputs, together with changing regulatory requirements across the global research environment, make SDTM transformation an increasingly burdensome task—and a major opportunity for AI-powered automation.

Parexel's SDTM tool: Connecting through standards and automation

Parexel's SDTM transformation tool is a generative AI-based solution, developed by Parexel and built on Palantir's AIP platform. AI automation eliminates manual processes by centralizing and integrating the multiple, disparate data sources and documentation involved in traditional SDTM transformation. It then automates the process of combining information from these documents and datasets to produce the study deliverables (i.e., automating these previously disparate and manual tasks done by SMEs) with a human in the loop, guided by AI, for review. Using foundational large language models in combination with complex algorithms orchestrated by AIP, Parexel's solution automates the creation of SDTM datasets, reducing programming time, accelerating data integration, and ensuring higher data quality. "We believe the combination of Palantir's AIP platform and Parexel's vision for clinical trial transformation will set new benchmarks for speed, accuracy and first-time quality," said Dan Ballard, Senior Vice President, Digital Enablement and Innovation. "Data standardization is the foundation for unlocking the value of AI in clinical research."

- **How it works.** At the beginning of a study, and before patient data are collected, SDTM mapping is automatically generated through ingestion of protocol content, EDC and non-EDC metadata, and SDTM specifications. Mapping is automatically updated throughout the study as protocols are amended or source data are modified.

  The generative AI-automated system consolidates the ingestion and abstraction of study documents and streamlines the workflow of developing SDTM deliverables, including datasets and submission-ready documentation (annotated CRF, for example). Users can leverage all the data components to inform trial operations when needed. For example, datasets generated from study documentation abstracts can inform trial design and monitoring approaches.

  The solution leverages ontologies to establish a well-defined model of data assets that can be systematically accessed and leveraged for downstream workflows such as CDISC controlled terminology and domain model objects to execute real-time data evaluation and perform ongoing health checks on ingested, transformed and exported datasets.

  The ontologies are surfaced to end users in intuitive and bespoke front-end applications to simplify the human oversight activities. They also enable users to provide direct input into generated content to ensure that the final submission deliverables are fit for purpose and are qualified for production use.

- **Real-time access on the way.** The platform will be further refined and optimized to provide real-time data access and operability. In the coming real-time paradigm, SDTM data will be generated within days of initial data ingestion and mapping. Access to earlier, reliable SDTM data will inform decisions at critical timepoints, such as interim analyses in adaptive trials and reviews by data monitoring committees. Earlier, higher-quality, standardized data opens up many possibilities for better reporting, cleaning, operational efficiency and faster development.

The benefits: Cutting timelines in half

Parexel's generative AI-enabled SDTM transformation platform is being deployed on select studies, with plans to scale throughout 2025. Based on experience to date, the AI-automated process reduces SDTM transformation timelines by approximately 50%; the time needed to produce SDTM deliverables is six weeks, compared to 12 weeks needed for the traditional methodology.

As the solution is refined, Parexel anticipates additional efficiencies that will enable SDTM transformation builds in only three to four weeks, with minimal effort required to refresh SDTM datasets based on updated source data.

- **Downstream impact.** Ultimately, SDTM automation is the critical foundational step for automating the entire data lifecycle in clinical trials, from initial data collection to regulatory submission. Earlier access to standardized SDTM data will accelerate and improve downstream operations. It will support earlier preparation of analysis programming and quality controls, for example. In pharmacovigilance, it will provide earlier data to support safety trend signaling and medical reviews. In the end, this translates to faster time to market for biopharmaceutical sponsors.
- **Optimizing operations.** Currently, we are using AIP with infrastructural and operational tools to streamline SDTM transformation and data delivery. The same frameworks are being used to optimize operations at scale. For example, although data conversion, analysis and dataflows are automated, the handoffs of these datasets to experts for review are usually manual. Manual handoffs slow the data delivery process and introduce risk for error. We are applying AIP to automate handovers, which drives more efficient and consistent collaborations.

AIP is helping to meet another management challenge—the scheduling and allocation of specialized resources over the course of a clinical trial. For a phase III trial, for example, the AI-enabled platform provides data to support resource management decisions involving hundreds of experts playing various roles in large, complex studies conducted worldwide.

> "Parexel is harnessing the full potential of AIP to hyperautomate mission-critical functions of Parexel's core business." said Ted Mabrey, Palantir's Head of Commercial. "The framework and rigor they've developed sets a model for applying generative AI to complex problems in a highly regulated space. We're excited to see how it continues to advance Parexel's vision of creating an AI-enabled clinical trial lifecycle."

The future of AI-enabled drug development

In every application, advanced AI technologies like large language models will free clinical researchers to focus on higher level tasks. Emerging AI + human-in-the-loop platforms will empower human experts to mine ever-greater sources of health data and use that intelligence to advance therapeutic innovation.

Early AI applications in SDTM transformation have shown significant benefits in automating this essential first step in the clinical data lifecycle.[3] Some current efforts are using AI technologies to create libraries of standards components—such as forms and fields for source data collection and standards rules—to improve efficiencies.

Parexel is the only CRO who is developing an end-to-end generative AI-enabled SDTM platform based in AIP. Built on a comprehensive data governance strategy and technology framework, this pioneering AI solution connects the data value chain to realize the potential for AI at scale and accelerate the delivery of new therapies to patients, serving as a model for the future of AI-enabled drug development. As this platform evolves, applications can go well beyond data standards to enable auto-generation of trial documentation. From study protocols, informed consent, database design and specifications to statistical analysis plans and clinical study reports, this future AI-powered system can fully connect and streamline the clinical trial data lifecycle.

References

1. S Abogunrin, JM Muir, C Zerbini and G Sarri, 2025. How much can we save by applying artificial intelligence in evidence synthesis? Results from a pragmatic review to quantify workload efficiency and cost savings. Front w,16:14542-14545. At: https://www.researchgate.net/publication/388569240_How_much_can_we_save

2. Palantir Technologies Inc. At: www.palantir.com

3. A Thukral, S Bhardwaj, 2023. The future of SDTM transformation: AI and HITL. Applied Clinical Trials. At: https://www.appliedclinicaltrialsonline.com/view/the-future-of-sdtm-transformation-ai-and-hitl

With Heart™

Parexel International Corporation
275 Grove St., Suite 101C, Newton, MA 02466
+1 617 454 9300
Offices across Europe, Asia, and the Americas
www.parexel.com

parexel®