Paper AD07

Enhancing Clinical Trial Management with Automated Data Integration and Real-Time Analytics

Katrijn Op De Beeck, Sopra Steria, Diegem, Belgium
Emmy Pahmer, OCS Consulting, ‘s-Hertogenbosch, Netherlands

ABSTRACT

Argenx is a biotech company managing multiple compounds across different phases of development, working with various contract research organizations (CROs). One key challenge has been achieving unified, timely oversight of operational performance and trial progress across portfolios. To address this, we developed a data mapping framework in SAS Life Science Analytics Framework (LSAF). Vendor data ingestion is set up in collaboration with each vendor (CRO) and the operational data is standardized internally. Another argenx team generates automated SDTM datasets. With both the clinical and operational data, we can generate important study metrics which are then used to create visualizations designed to meet the needs of our internal stakeholders. This fully automated integration enables daily-refreshed, cross-study and cross-vendor analytics, driving better visibility into study progress, vendor performance, and protocol compliance. We will share how this architecture was developed, the tools and processes we employed, and how our visualization strategy enables meaningful insights across our clinical programs.

INTRODUCTION

In today’s increasingly complex clinical trial landscape, the ability to make timely, data-driven decisions is a necessity. At argenx, a biotechnology company committed to innovation and patient-centricity, the exponential growth of outsourced clinical trials has introduced new challenges in operational oversight, data harmonization, and real-time visibility. To address these, argenx has embarked on a journey to automate the ingestion of clinical operations data from external vendors.

The traditional model - where each trial manager tracks study progress via the CRO clinical trial management system (CTMS) system - proved insufficient for the scale and speed required in modern trial execution. Data latency, inconsistencies across vendor systems, and limited interoperability hindered the ability to monitor study progress across compounds and indications. These limitations not only delay insights but also compromise the agility needed to respond to emerging risks and opportunities.

The Operational Insights Project at argenx is a strategic initiative designed to enhance clinical trial oversight by enabling near real-time access to key operational metrics, including study progress, site activation, and participant enrollment. This initiative addresses the growing complexity of outsourced clinical trial models by establishing a standardized and automated data integration pipeline across multiple vendor systems.

METHODOLOGY

The architecture for the Operational Insights initiative was designed to address the challenges of fragmented data sources in a fully outsourced clinical trial model.

DATA INGESTION – OPERATIONAL DATA

At argenx, operational data originates from multiple CROs, each with distinct systems and formats. Each vendor has its own CTMS system for tracking study progress and important milestones. To harmonize this landscape, we developed a standardized data mapping framework within the SAS Life Science Analytics Framework (LSAF). This framework enables the transformation of vendor-specific CTMS data into a unified internal schema, supporting consistent analytics across studies and vendors. Data ingestion onto SAS LSAF is executed via daily secure sFTP transfers or direct Snowflake connections, depending on the vendor’s infrastructure. The CRO data is shared as CSV, TXT or SAS files. Once ingested, this data is mapped to internally defined standards through SAS programming in LSAF.

DATA INGESTION – SDTM DATA

Argenx started a project called 'Near Real-Time Data Flow' almost 3 years ago. There was a high need from multiple stakeholders in Clinical Development to get access to near real-time clinical study data.
Monthly clinical SDTM data received from DM vendors was no longer sufficient as a fast growing biotech. The project started with pulling in nightly RAVE EDC (eCRF) data from our different Rave URLs into SAS LSAF; later on the project was expanded to include Central Lab Partners. Then automated jobs were developed within argenx to create SDTM datasets. This is called generic SDTM (gSDTM), and is intended for internal analytics.
In this project, gSDTM data is generally used for "actual" numbers, such as number of screen failures, number of subjects discontinued, etc. while CTMS data is used for planned recruitment and various other milestones.

DATA INGESTION – OTHER SOURCES

While the long-term objective of the Operational Insights initiative is to achieve fully automated ingestion of operational data from all contracted CROs, interim solutions remain necessary. In cases where automated data feeds are not yet available or vendor systems lack sufficient granularity, argenx continues to rely on manually created Excel files. These files are securely stored in SharePoint and programmatically ingested into SAS LSAF, where they are processed alongside automated data streams. This hybrid approach ensures continuity in operational oversight while maintaining flexibility during the transition to full automation.

PROGRAMMING

The programs used to perform the mapping are mostly based on macros which are run depending on the source of the data. Incoming data files differ significantly in format, column naming conventions, and structure. These heterogeneous inputs are transformed into standardized datasets in accordance with predefined technical specifications. For each use case, a dedicated dataset is generated per study. These individual study-level datasets are then aggregated to produce a consolidated output that includes all studies for that specific use case. This approach ensures both granularity and scalability in reporting. At the moment there are over 70 studies in this project, reflecting the broad adoption and operational relevance of the solution.

AUTOMATION

A cornerstone of the Operational Insights initiative is the implementation of a fully automated workflow within SAS LSAF. This automated process flow is scheduled to execute daily and orchestrates the end-to-end process from data ingestion to pushing data to the external environments. These outputs serve as the foundation for dynamic visualizations in Tableau and Elluminate, enabling near real-time oversight across the clinical portfolio.

USE CASE IDENTIFICATION AND MAPPING IN SAS LSAF

Use cases typically originate from operational pain points or strategic reporting needs identified by stakeholders in Clinical Trial Operations, Data Acquisition, and Medical Affairs. These needs are discussed in cross-functional workshops and documented in initial business requirement assessments. Once a use case is defined, a corresponding User Specification document is created. This document outlines the business context, objectives, data sources, mapping logic, and expected outputs. It serves as a bridge between operational needs and technical implementation. The specifications are authored collaboratively by Clinical Operations Systems Specialists and Data Integration Programmers.

USE CASES

Some examples of use cases that have been developed under the Operational Insights initiative:

Across Study Dashboard: This dashboard provides a cross-trial view of key operational metrics such as site activation, enrollment information, and study milestones. It integrates data from CRO CTMS, SDTM, and manually created Excel files, enabling near real-time oversight across the portfolio.

Recruitment Rate Dashboard: This dashboard visualizes recruitment performance at study, country and site level. It compares planned versus actual enrollment and highlights underperforming regions for targeted intervention.

Delta PI Report: This report was designed to support regulatory compliance with 21 CFR 312, which mandates that sponsors notify the FDA within 30 days of adding a new investigator to a clinical trial protocol for U.S. sites. In addition to regulatory obligations, the report also addresses business needs related to site and patient engagement, specifically the timely publication of activated sites on ClinicalTrials.gov, which requires confirmation of site activation prior to listing.

Protocol Deviation Dashboard: this dashboard is a visual interface for monitoring subject-level protocol deviations across clinical trials. It is designed to support operational oversight through interactive dashboards.

DATA OUTPUT DESTINATIONS AND SYSTEM INTEGRATION

Standardized datasets generated for each use case are programmatically pushed to designated output environments based on their intended use, compliance requirements, and stakeholder access needs. Two primary platforms are utilized: Snowflake and Elluminate.

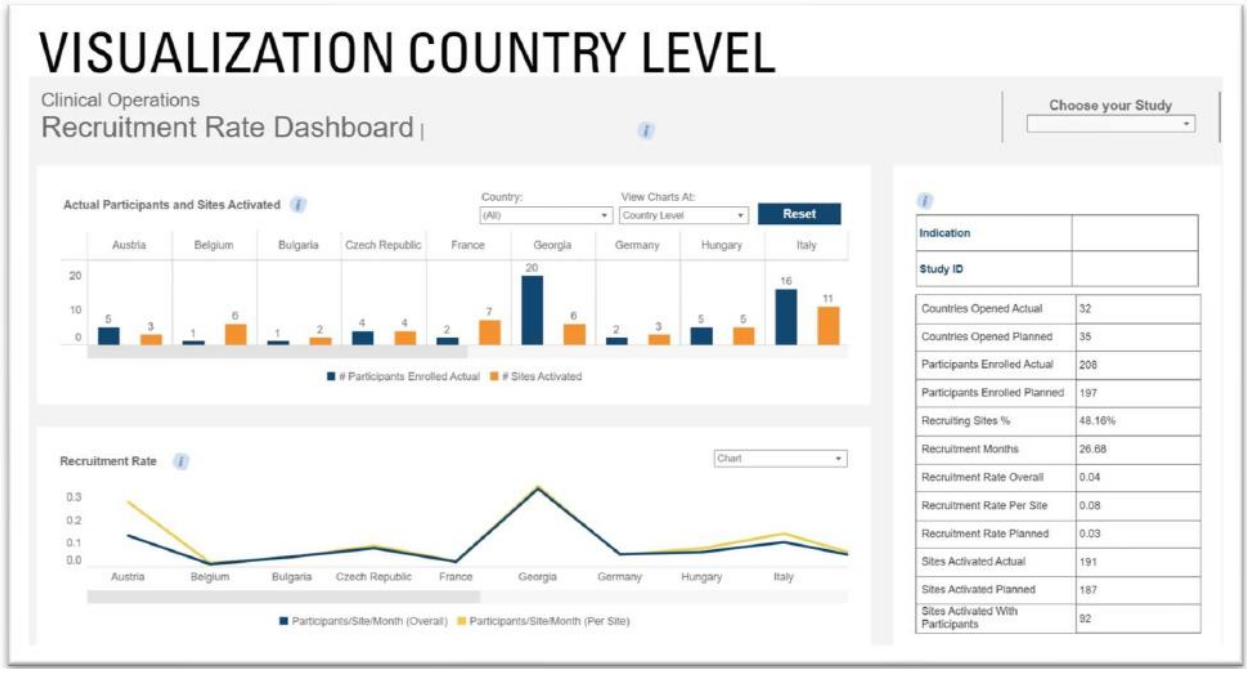
Snowflake serves as a cloud-based data warehouse that supports scalable storage, processing, and sharing of operational datasets. It is primarily used for non-GxP compliant reporting and visualization. Data pushed to Snowflake is consumed by Tableau Cloud, where interactive dashboards are developed to support portfolio-level oversight, recruitment tracking, and milestone adherence. These dashboards are refreshed daily and are accessible to internal stakeholders for strategic planning and performance monitoring.

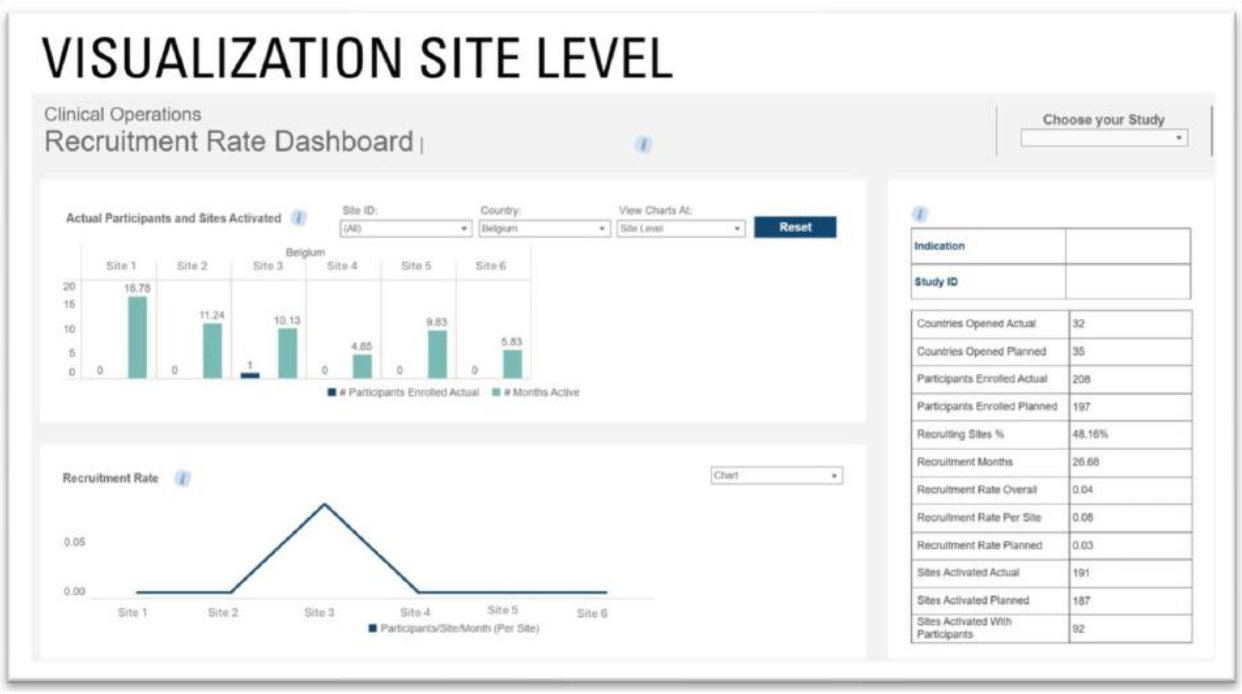
Elluminate is a cloud-based platform that integrates and analyzes data for clinical trials. Elluminate has multiple systems for reviewing data and is GxP compliant. Elluminate is used for our use cases because it supports audit trails and role-based access, and regulatory-grade traceability. It enables cross-functional review by Clinical Operations, Data Management, and Franchise Leads.

This dual-platform strategy allows argenx to balance flexibility and compliance, ensuring that each use case is supported by the most appropriate technical environment while maintaining data integrity and operational efficiency.

VISUALIZATIONS

TABLEAU: Recruitment Rate Dashboard





STAKEHOLDERS/END USERS

Clinical Trial Managers are among the primary users of the dashboards and reports generated through the initiative. They utilize real-time metrics on site activation, enrollment progress, and milestone adherence to monitor study execution and proactively address operational risks.

Medical Monitors benefit from integrated views of subject-level data, particularly through modules such as Protocol Deviations (PD) in Elluminate CTOA. These visualizations allow them to identify trends in deviation categories.

Upper Management leverages high-level portfolio dashboards in Tableau Cloud to track strategic indicators such as recruitment rate, protocol adherence, and vendor performance.

Regulatory Liaisons use outputs such as the Delta PI Report to ensure timely compliance with regulatory requirements, including FDA notification of newly added investigators and site listings on ClinicalTrials.gov. The automated generation and distribution of these reports streamline workflows and reduce the risk of non-compliance.

CHALLENGES

While the Operational Insights initiative has significantly advanced the automation and centralization of clinical operations data, several challenges still exist. One of the most persistent obstacles is the retrieval of CTMS data from external vendors. Not all CROs provide automated, structured data feeds, and even among those that do, there is considerable variability in data formats, naming conventions, and update frequencies. This variability complicates the standardization process and sometimes requires study-specific mapping logic and ongoing vendor engagement.
For legacy studies, where automated CTMS feeds are unavailable, argenx relies on CRO SDTM data. These datasets are often inconsistent across studies, particularly within the Disposition (DS) domain.
Additionally, data availability is a limiting factor for studies in early startup phases. In such cases, sites may be activated, but participant-level data is not yet available, resulting in incomplete datasets and delayed dashboard implementation.

NEXT STEPS

As the Operational Insights initiative continues to evolve, several strategic next steps have been identified to broaden its impact and enhance its utility across clinical operations. A key priority is further collaboration with CROs to improve access to operational data. While several vendors currently provide structured CTMS feeds, others still rely on submitting data manually or are unable to provide data in the required detail. Strengthening partnerships and aligning on data standards will be essential to expand coverage and reduce reliance on manual Excel files.

New use cases are actively being discussed and scoped for future dashboard development within the Operational Insights framework.

CONCLUSION

The Operational Insights initiative at argenx represents a transformative step toward modernizing clinical trial oversight through automation, standardization and real-time analytics. By integrating operational and clinical data from diverse vendor systems into a unified analytics framework, the project has enabled scalable, near real-time visibility across more than 70 studies. This has empowered stakeholders to make faster, data-driven decisions that improve trial efficiency, quality, and compliance.

CONTACT INFORMATION

Your comments and questions are valued and encouraged. Contact the authors at:

Emmy Pahmer, Clinical Data Analyst
OCS Consulting for argenx
emmy.pahmer@ocs-consulting.com
https://ocs-consulting.eu/

Katrijn Op De Beeck
Sopra Steria for argenx
katrijn.opdebeeck@soprasteria.com