

APPLIED CLINICAL TRIALS

YOUR PEER-REVIEWED GUIDE TO GLOBAL CLINICAL TRIALS MANAGEMENT



CLINICAL DATA MANAGEMENT



DATA MANAGER ROLE

FUTURE ROADMAP

DATA SCIENCE INVESTMENT

UNLOCKING VALUE

CLINICAL TECHNOLOGY

Solving the Data
Scalability Problem

TRIAL INSIGHTS

Vendor Qualification:
Process Improvements

CLOSING THOUGHT

Keeping Up With Data
Management Changes

Scalability: Solving the Unique Clinical Trial Data Problem

Keith Aumiller

Finding a middle-ground approach to balancing new solutions arising from data science with traditional requirements for data collection and submission.

There is a problem in clinical trial data structures. It isn't a problem for any individual trial, or even a few. But, for those of us who deal with tens or hundreds or even thousands of trials worth of data, it is a pretty serious hurdle. It's called scalability, and for the clinical trials industry, it presents a unique issue.

The reason for this is the expectation in clinical trials that data be maintained in a separate compartmentalized database. This is considered good clinical practice (GCP). Decades ago, when this practice was first introduced, it set a series of goals that were admirable, expected, and not difficult to follow. However, as the years went on, the rules were not updated. Much like trying to drive a modern car on streets designed for horse and carriage, what was once a good fit for the industry has become somewhat of a hindrance.

Simply put, a lot has changed technologically since the original principles were put in place decades ago. The tools we have at our disposal are on a level that the original guidelines could never have predicted. This means that the expectations for the use of data in general, not only in clinical trials, couldn't have been anticipated.

So, how do we marry the expectations and solutions we have available to us from a data science perspective with requirements arising from these traditional viewpoints? Well, as an industry, we must recognize the limitations of the physical models we have in place currently. Then, we must talk about the possibilities that are available to us. Fortunately, other industries have solved these problems and are making advancements. So, while we are certainly lagging behind, there is a clear light at the end of the tunnel.

In order to begin advancing, we must first understand the exact problem. As mentioned, the problem that we

must solve is that all clinical trial databases must be kept separate from each other. The debate is between physical separation and logical separation.

Logical separation provides the ability to be in one electronic database but to utilize filters to isolate the clinical trial database that is relevant for your analysis.

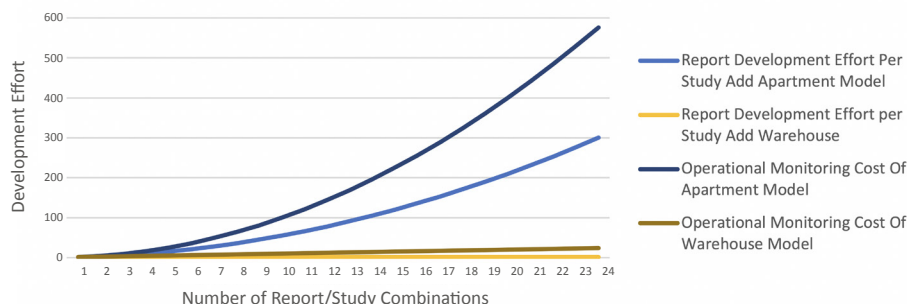
As an industry, we must recognize the limitations of the physical models we have in place currently. Then, we must talk about the possibilities that are available to us.

Physical separation puts major hurdles around analysis and automation, while providing an additional layer of security and integrity protection. Both strategies are valid, but only one of them is scalable operationally, and allows for automation of data science, analysis, and monitoring.

- **Physical separation** provides major technological hurdles. This means that data normalization and standardization can't be enforced with integrity controls. Physical separation requires separate user identification management and controls at the schema level of the database and adds a layer of overhead to the implementation that can be cost prohibitive when implementing reporting and operational solutions at scale (see Figure 1 on facing page).
- **Logical separation** also affords the ability to undergo development to a "comprehensive" re-

CLINICAL
TECHNOLOGY

Scalability Issues



Source: Aumiller

Figure 1. Physical separation requires separate user identification management and controls at the schema level of the database and adds a layer of overhead to the implementation that can be cost prohibitive when implementing reporting and operational solutions at scale.

menting security controls at the record level of the schema.

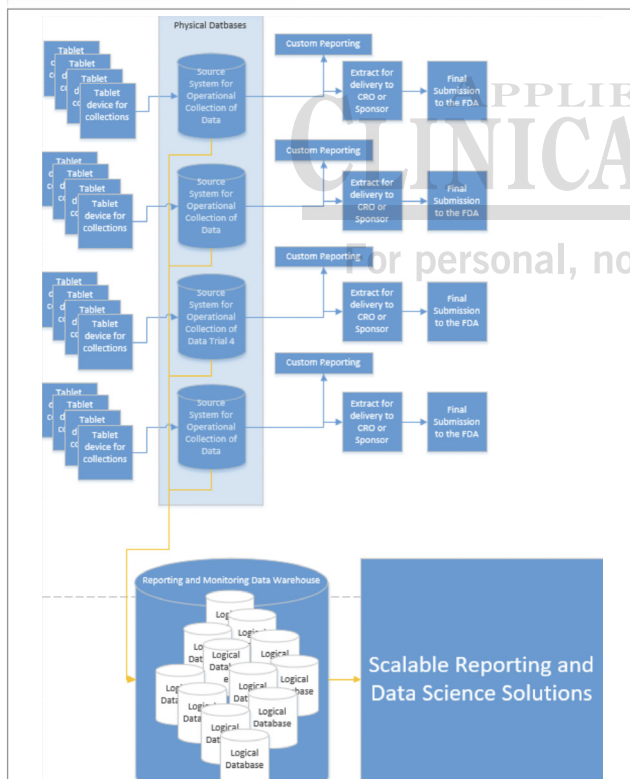
A nice middle ground to this is the apartment model. Here, the clinical submission data flow path works in a physical separation “apartment” model while having a system that isn’t on the critical path of collection and submission house data in a warehouse or data lake for scalable reporting, operational monitoring, and aggregated data analysis, and aggregated data science statistical model development (see Figure 2).

This allows data collections to meet all previous regulatory concern requirements, get the most out of modern data architecture and data science solutions, and does so without ever compromising GCP principles.

out ever compromising GCP principles.

If one wants to leverage image analysis, classification, machine learning, deep learning, or any of the other potentially groundbreaking

Apartment Model



Source: Aumiller

Figure 2. Here, the clinical submission data flow path works in a physical separation “apartment” model.

If one wants to leverage image analysis, classification, machine learning, deep learning, or any of the other potentially groundbreaking technologies that are available on specific data, The data must be in a structure that normalizes and standardizes that data.

ing technologies that are available on specific data, the data must be in a structure that normalizes and standardizes that data.

In other words, in order to get the most out of technology, while also staying within GCP, one will need to marry these two architecture solutions together and develop a comprehensive answer to these problems.

As we move further into the decade, we can expect to see some groundbreaking and new ways to use technology, both in our personal lives and in clinical trials. The question for us as an industry, however, is will we be ready when those advancements arrive? Or, will we still be stuck driving supercars on top of cobblestone?

porting and data science solution. While logical separation has risks associated with security and access control, those can be mitigated with access control groups and even imple-

Keith Aumiller, Senior Director, Data Services, Signant Health