MEDIDATA | Rave

WHITE PAPER

# Modernizing Clinical Data Management to Be Scalable, Flexible, and Intelligent

# Table of Contents


| | |
|---|---|
| Introduction | 3 |
| The Mounting Data Pressures of Modern Clinical Trials | 4 |
| Transforming CDM for Today's and Tomorrow's Trials | 6 |
| The Three Pillars for Intelligent CDM | 8 |
| Skill Sets Need to Evolve Too | 11 |
| Summary | 12 |
| References | 13 |

# Introduction

Clinical trials are becoming more complex and a confluence of forces—including the ongoing COVID-19 pandemic, greater emphasis on patient experience, increasing protocol complexity, and precise value differentiation—is accelerating this complexity. Trials have evolved to include an ever-expanding array of data sources, increased data volume and precision, decentralized clinical trials (DCT), and adaptive designs.

These environmental and industry changes have led to significant data management challenges because clinical data management (CDM) technologies and processes have not progressed at the same breakneck speed. Further, technology ecosystems are often built with a mix of disparate tools that are homegrown or from different vendors, and the individual components lack interoperability. Consequently, trial data is largely managed with tools and processes that are not able to evolve and adapt at the speed required to support the realities of the ever-changing modern clinical trial.

The time for advancement in CDM is now. It is currently being modernized by unifying data and workflows on a cloud-based platform. This is transforming CDM to be scalable, flexible, intelligent, and interoperable with a variety of third-party systems so that it can sustain drug development into the future. That future includes more DCT designs where visits are monitored remotely, drugs are shipped to patients, data is captured directly from patients regardless of location, and all data cleaning and transformation occurs in real-time using intelligent and automated digital oversight tools remotely.

The benefits of unifying data and workflows on a cloud-based platform include an aggregated view of patient data, elimination of duplicate data entry and reconciliation, and streamlined processes. In short, a unified platform perspective provides the necessary framework needed to achieve the speed and scale required of modern trial designs including:

- Faster study start-ups (through use and compliance management of reusable libraries)
- Far more efficient study conduct
- Streamlined database locks

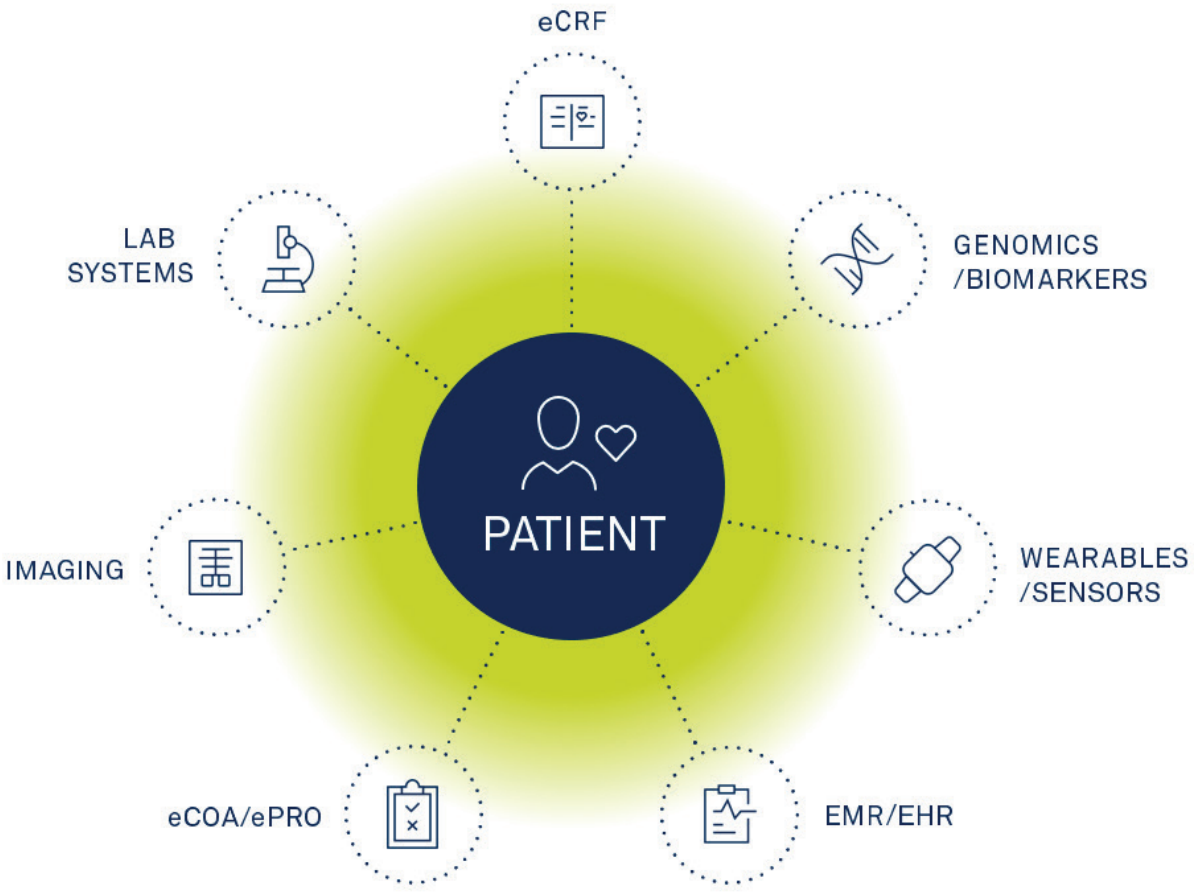
Ultimately, unified platform solutions can boost the value differentiation of your new drug or device by developing a more comprehensive profile for how patients are affected by drugs and disease. The importance of this is heightened in today's environment where stakeholders across the healthcare enterprise—regulators, payers, and patients—are placing greater emphasis on patient-centricity and value demonstration of new therapies.

This white paper provides a brief summary of why CDM must quickly adapt to the mounting data pressures in modern clinical trials. Next, it discusses the three pillars that form the foundation of a modern intelligent CDM platform that is needed to succeed in an increasingly complex clinical trial world. Lastly, two case studies are provided; the first one highlights how missing data in a vaccine mega-trial was detected in near real-time via intelligent, automated data-review technology, and the second describes how negative patient outcomes were averted by taking a holistic approach to patient data capture and monitoring.

# The Mounting Data Pressures of Modern Clinical Trials

Clinical trials have mushroomed with highly sophisticated protocols and multiple divergent data sources, inordinate data volume, decentralization, and increasingly complex adaptive designs. One estimate indicates that data points collected in Phase III trials have increased threefold in the last 10 years, to an average of 3.6 million, and increased trial complexity is driving the expanding volume arising from an increasingly wider array of sources (Tufts CSDD, 2021). Since data first began to be electronically collected using eCRFs (electronic Case Report Forms), the data acquisition landscape has evolved greatly (Figure 1). Data is now being collected by many different modalities, and the technology in each one has rapidly improved, meaning these new forms of data capture and utilization have increased at a very high rate, with eCRFs now being used to capture only ~30% of clinical trial data; a value that is expected to further decrease as more DCTs come online and a smaller percentage of data will be collected via eCRFs.

**Figure 1: The growth of clinical data is driven by increased volume arising from a growing array of sources.**



This trend is only accelerating, given the pandemic and greater emphasis on patient-centric approaches over time. The former has boosted the demand for DCTs and virtualization within trials, leading to a sharp rise in the adoption of digital health technologies (DHTs), including sensors and wearables, further accelerating the available types of high-volume, high-velocity, high-precision measurements. This is anticipated to increase, with 70% of trials projected to use DHTs in the next few years alone (Myshko, 2019) to capture both objective (e.g., sensors/ wearables) and subjective (e.g., ePRO) data streams that complement each other and significantly enhance evidence generation. Together, the depth and breadth of data from all of these sources have challenged the industry as to how they should and could be used outside the traditional data cleaning and reconciliation activities.

As mentioned, CDM—which includes how data is reviewed, cleaned, and locked—technology and processes have not progressed at the same pace as innovations in data capture. Currently, multiple data sources are aggregated in one of three ways:

- They are forced into electronic data capture (EDC) systems to leverage review and cleaning tools only available in those siloed systems. This comes with a management overhead within the EDC system, that is not related to the eCRF, and often interferes with the site data capture for which they are designed.

- They are manually consolidated by programmers with issues managed in spreadsheets, emails, and other highly manual processes with little to no tracking mechanisms.

- Organizations develop their own bespoke solutions to aggregate and manage data. These come at an extremely high development cost, require constant maintenance, and struggle to keep up with the changing world of data acquisition. Additionally, they are not the true source of the data and they do not support interoperability with systems outside the organization.

All of these aggregation mechanisms are brittle and unscalable, use a highly manual process that is often error-prone, and have an unacceptable degree of latency.

Further, several lines of evidence have raised questions regarding the effectiveness of traditional data review processes; all suggesting that data changes very little after initial entry.

For instance, one analysis showed that, on average, only 3.7% of data entered in eCRFs are changed after entry in EDC systems, with only 2.6% of changes not attributed to source data verification (Sheetz, 2014). Additionally, Stokman et al. (2021) reported on their analysis of studies comprising more than 20,000 participants, nearly 50 million data points, and more than 1.9 million queries. They found that while the overall query rate was 3.9% (including indirect and noninformative modifications), fewer than half of these queries resulted in a data change, affecting less than 1.7% of entered data. Given the limited impact of the tremendous efforts that go into generalized query-based data cleaning, the article concluded that it does not contribute proportionally to the quality of the final database used for analysis. The authors, therefore, recommend ending the current query process to correct errors of noncritical data in Phase III studies and placing greater emphasis on tools and techniques that help identify systemic issues in the data collection process.

Modern CDM platform solutions must be able to easily surface and transform data for visualization and analyses. This includes seamlessly combining discrete, highly diverse, and even historical datasets, allowing for like-to-like comparisons between different data models to unlock the full value from diverse data streams.

The following section summarizes the key elements for modernizing CDM to meet the trial demands of today and tomorrow.

# Transforming CDM for Today's and Tomorrow's Trials

On the road to modernizing CDM, a holistic aggregation of data is required. This will allow for a complete study and patient dataset that has the context of the whole study design. This requires intelligent and automated tools and workflows and a user experience that can operate on the complete study and patient dataset agnostic to its source (Figure 2).

Examples of these intelligent, automated tools and workflows include the following:

- **Centralized study design and data.** Enriching data with semantic meaning provides several benefits including automated common data review to eliminate duplicative tasks, enhanced data utilization with a single source of truth for clinical data and study design, automated data cleansing and reconciliation processes, and real-time data transformation to standardize submission-ready outputs.

- **Auditable, permission-driven issue management capability on all of your data.** Edit and consistency checks are configured across data sources using an "EDC-like" rules engine. This reduces duplicative tasks across study execution and management stakeholders and allows other stakeholders to be added, such as central lab providers, who have traditionally been notified of reconciliation errors through inefficient processes that rely on emails and spreadsheets.

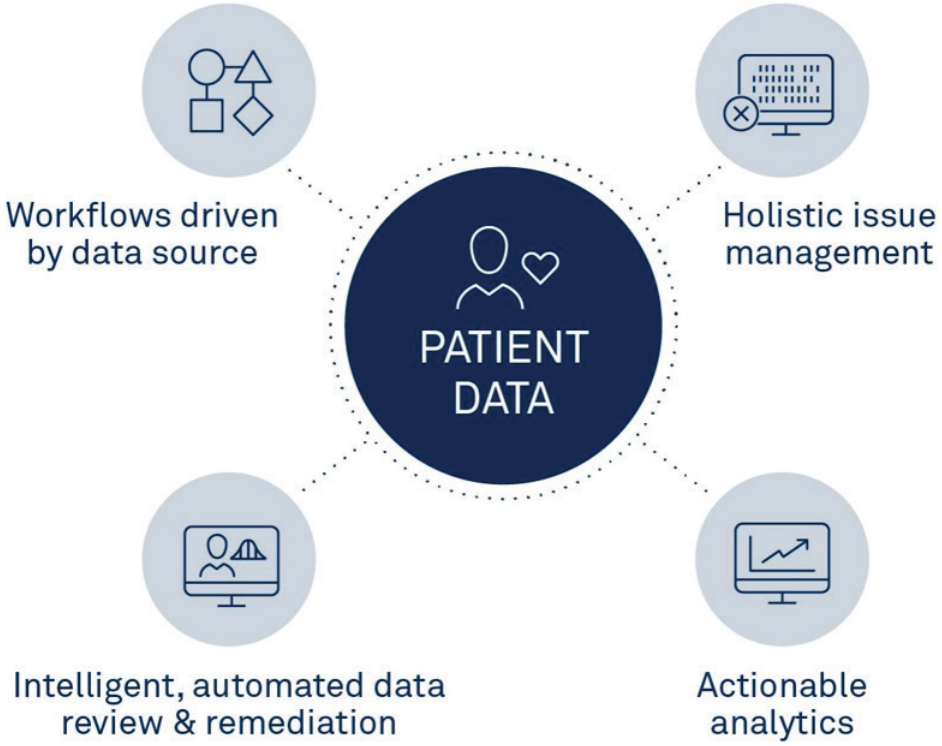
- **"Analytics first" approach to drive and direct users to action.** This approach provides stakeholders with a clear view of the study and overall program health, which can be tailored based on specific areas of interest. This capability also provides a highly collaborative environment that reduces duplicative tasks.

- **Intelligent approach to data review.** Traditional data review processes are not easily scalable to adjust to the increasing volume of new data sources and data volume. Intelligent approaches to review can highlight outliers, patterns, and inconsistencies without human intervention while reducing site burden by using electronic health record (EHR) or electronic medical record (EMR) data sources to eliminate duplicate entries.

- **Data source-driven workflows and activities.** Single static workflows are becoming insufficient to meet the fluid demands of modern trials. Workflows and activities should be determined by their data source and means of capture. Currently, the same data in a study design is often acquired using a variety of mechanisms within a singular workflow. This has been modernized so that required activities are automatically driven by the data source; this is accomplished with algorithms to detect anomalies, inconsistencies, and patterns across all data sources.

- **Predictive approach to resource allocation.** Tools traditionally used to manage the study resources are often inefficient and they do not always take a comprehensive view of all variables that can predict workload. A data-driven approach to managing resources is achievable by combining historical metrics with study behavior and projections of how much data will be collected and processed.

**Figure 2: Modernizing CDM demands holistically aggregating data into a complete study and patient dataset that has the context of the WHOLE study design. This requires intelligent and automated tools and workflows and a user experience that can operate on the complete study and patient dataset, agnostic of its source.**



# The Three Pillars for Intelligent CDM

- ✓ **Unified 360° view of patient data** aggregated from multiple sources in real-time and augmented with semantics to provide a deep level of context to the study design,
- ✓ **Risk-based data management** with near real-time, automated issue detection using machine learning algorithms, and
- ✓ **Automated data transformation** for analysis, interoperability (e.g., cross-study analyses) and reuse (e.g., synthetic control arms).

## UNIFIED 360° VIEW OF PATIENT DATA

By merging disparate data streams through a unified, intelligent, and secure platform that is designed to process data through a seamless common model, you can now easily and quickly visualize and analyze data in a unified way to drive the value demonstration of your new drug or device. Many of these data streams are complementary and uniting them means that you can develop highly comprehensive patient profiles that are more informative with regards to how patients are affected by drugs and diseases. Here are some of the key features enabling this:

**Agility:** Study designs must be more data acquisition focused, regardless of source across all data assets. Data capture and management processes must be designed with agility in mind to enable changes in data capture to meet the custom needs of an individual study, site, or patient (e.g., ePRO, eConsent, sensors/wearables). It is important to define the full schedule of assessments (visits), the set of activities of the protocol, and any streaming or other study aspects that may not be visit related. This will also include other characteristics, including cohort, arm, and study phase design, that provide context and drive behavior during execution.

**Common Data Model:** Must support a wide variety of data from a diverse group of sources and co-localize the disparate data for use in algorithmic cleaning, AI-assisted workflows, and custom data extractions and models. For instance, whether "systolic blood pressure" is designated as SBP or Sys_BP does NOT drive the meaning of the data point; rather, the semantic meaning, which is systolic blood pressure, is the driver of the meaning. A common data model allows drug developers to stay focused on the implications of their analyses rather than expending time and resources on the nuances of ingesting and harmonizing disparate datasets.

**Scalable:** Data capture must be scalable for all types and volumes, ranging from simple laboratory test results to high-volume, high-velocity data from medical-grade DHTs. Scalability is also achieved by automating data review with machine learning and displaying downstream outcomes in easily interpreted visualizations based on a single source of truth that allows all program and study stakeholders to consume the data's meaning.

**Secure:** When asked to rate EDC product attributes by their impact on clinical trial success, respondents to a 2020 Industry Standard Research survey selected data security as the attribute with overall highest impact, with 48% and 34% saying it has high or moderate impact, respectively (Industry Standard Research, 2020). Therefore, you should seek a partner that has a solid track record of data security and acting as a custodian of patient trust. They should have ISO 27701:2019 certification, which is the closest to a General Data Protection Regulation certification that can be obtained, and SOC 2 Type 2 "Plus," which has additional controls, over and above the basic AICPA professional standards for both the Information Security and Privacy trust principles.

## RISK-BASED DATA MANAGEMENT

An estimated 4% of data does not change after entry, while 80% of edit checks are not triggered. Modern risk-based data management focuses on high-impact data points and relies on intelligent analytics (Figure 3). This is because every data point is not equal, and those with the greatest impact should be prioritized. Proactive, continuous review in almost real-time allows issues to be resolved before they become a systemic problem that can stifle trial progress and delay database lock.

> **"It's okay not to be perfect. Every data point is not equal, and we should prioritize data points that have the highest impact."**
>
> Wayne Walker, SVP Product, Rave Platform Technology (Medidata)

A risk-based data management approach has been shown to reduce edit checks by 20%–40% and allows for automating 50%–55% of data reviews. While traditional data review, cleaning, and reconciliation practices have meant that the final database lock has taken a number of weeks, we believe modernized CDM can accomplish this in hours or days. Ongoing intelligent and incremental task completion leaves fewer tasks to complete after last patient, last visit.

**Figure 3: Risk-Based Approach for Modern CDM. Traditional Methods Do Not Scale.**


## CASE STUDY: IDENTIFYING DATA ISSUES IN VERY LARGE DATA SETS

In a recent "mega-trial" designed to investigate the safety and efficacy of a novel anti-SARS-CoV-2 vaccine, an intelligent and automated analytical platform was deployed to detect under-reporting of adverse events (AEs) in almost real-time. Despite the mega-trial's rapid enrollment and high volume and velocity of data collection, the technology successfully identified missing AE data. The technology triggered a signal where deeper analysis showed that missing AE data was due to incomplete eCOA diary entries. Since the missing data were identified in almost real-time, they were quickly rectified and the sponsor was able to avoid compromising the statistical power of the study.

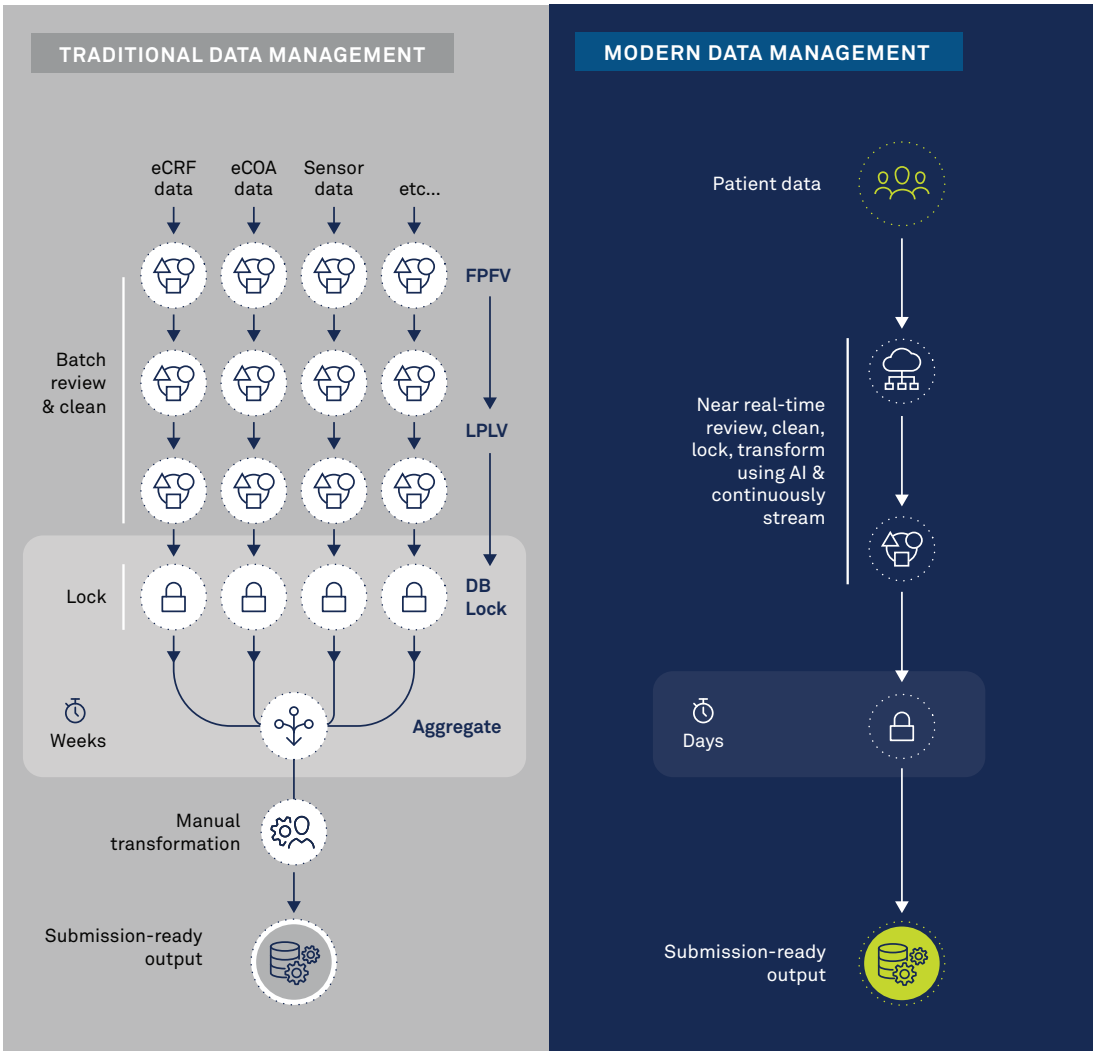
Due to the sheer size of the trial, combined with the volume and velocity of data collection, traditional CDM processes would not have been able to support the real-time review and analysis that was required by this important trial.

## AUTOMATED DATA TRANSFORMATION

There are multiple data format standards used at different stages of clinical trials, and after database lock, data in different systems usually must be aggregated and transformed. That can take a significant amount of effort and time, typically weeks. However, with increasing eSourced data, database locks can occur much quicker, since the data can be processed in intelligent and automated ways, eliminating the need for manual monitoring and source verification.

This is possible using modern CDM platforms that seamlessly aggregate data in real-time and process and inspect the data via remote digital tools with automated and intelligent algorithms to detect anomalies, inconsistencies, and patterns that should be questioned across all data sources (eCRF, ePRO, Sensors, eSource, etc.). Figure 4 provides an illustrated comparison of the traditional and modernized CDM processes.

**Figure 4: A Comparison of Traditional Versus Modernized Data Management Processes.**



Semantic meaning of data is then leveraged to automate transformation to standardized, submission-ready output. Rather than batch transformation, data is streamed from source to the endpoint within the transformed output. Transformation can be supplemented with mapping activity that can repeat transformations where semantics have not met the exact need.

## Skill Sets Need to Evolve Too

Clinical data management teams need to evolve their skill sets to adopt risk-based data management and leverage the advanced analytics and visualizations that platform technologies can provide. So as CDM evolves, data managers "will need advanced competencies to generate the high quality and high integrity data needed to drive the expected study outcome" (SCDM Reflection Paper, 2020).

### EXAMPLE SCENARIO: HOLISTIC APPROACH TO PATIENT DATA CAPTURE AND MONITORING

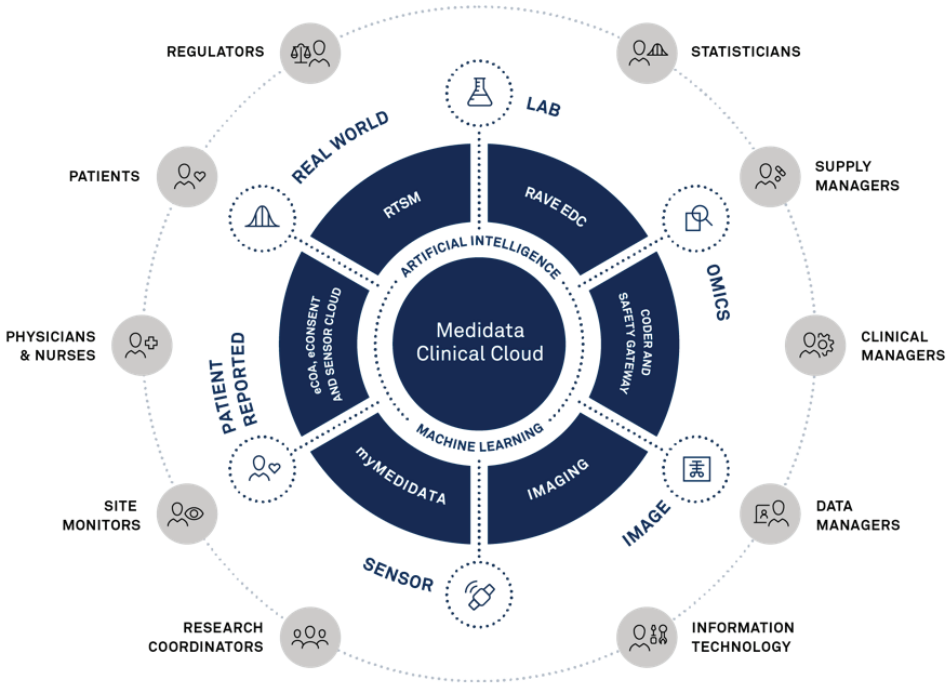
In a clinical trial conducted to test a new type of diuretic in patients with congestive heart failure, a holistic approach to patient data capture and monitoring helped avert negative patient outcomes. The trial involved remotely collecting data using various sources, including a biosensor patch, a smart weight scale, smart urine strips, eConsent, and ePRO.

Since the site was able to view both the sensor and ePRO data, they were able to quickly detect data anomalies that they could then act on, such as determining if patients needed to be checked on by a nurse via video visit and whether the patients should go to a hospital for additional tests. By using this comprehensive approach, negative patient outcomes were averted due to real time data observed by multiple stakeholders (site, sponsor, and CRO), including the ability to review sensor and ePRO data together, and withdrawing patients if necessary.

# Summary

The future of CDM is here, and it is scalable, flexible, interoperable, unified, and intelligent to support today's and tomorrow's clinical trials, which are increasingly reliant on the ability to synthesize high-velocity, high-volume patient-centric data into action-oriented, real-time insights and ultimately drive better study execution models. Modernized CDM platforms are enabling study constructs to be defined across all data sources; all data streams are consolidated into a single unified CDM platform, with data source agnostic tools that are intelligent, drive streamlined and automated workflows, and facilitate interaction between all stakeholders in a trial, all while driving insights that are not possible using manual approaches to CDM (Figure 5).

**Figure 5: Unified Clinical Data Capture and Management.**




This scalable, flexible, and intelligent approach is enabling faster study start times, faster database lock, higher data quality, and reduced effort for data review, clean, lock, and transformation. Ultimately, these solutions are allowing for more efficient and comprehensive value demonstration to increase patient accessibility and market differentiation.

*Medidata pioneered EDC, and has remained on the cutting edge of CDM with expertise, experience and data spanning thousands of clinical studies. Medidata is now modernizing CDM with a unified cloud-based platform powered by unique, automated, and intelligent processes, all supported by a highly experienced team. By partnering with our clients to develop the best CDM solution tailored to their needs, they are now unlocking key insights from their clinical trial data which is providing maximum value differentiation of their new therapies. For more information, visit [https://www.medidata.com/en/clinical-trial-products/clinical-data-management](https://www.medidata.com/en/clinical-trial-products/clinical-data-management)*

# References

Industry Standard Research. EDC Market Dynamics and Service Provider Performance Report. December 2020.

Myshko, D. Wearables in Clinical Trials. PharmaVoice.com. March 2019. Available at: https://www.pharmavoice.com/article/2019-03-wearables/

SCDM. August 2020. The Evolution of Clinical Data Management to Clinical Data Science (Part 3: The evolution of the CDM role) – A Reflection Paper on the evolution of CDM skillsets and competencies.

Sheetz N, Wilson B, Benedict J, et al. Evaluating Source Data Verification as a Quality Control Measure in Clinical Trials. Therapeutic Innovation & Regulatory Science. 2014;48(6):671-680.

Stokman P, Ensign L, Langeneckhardt D, et al. Risk-based Quality Management in CDM An inquiry into the value of generalized query-based data cleaning. Journal of the Society for Clinical Data Management 2021;1(1).

Tufts Center for the Study of Drug Development. 2021. January/February Tufts CSDD Impact Report: Rising Protocol Design Complexity is Driving Rapid Growth in Clinical Trial Data Volume. January/February Vol. 23 No. 1.