Clinical Data Workbenches: A Buyer's Guide

Veeva

Today, clinical trials depend on increasing volumes of patient data from more diverse sources than ever before.

An EDC system produces only about 30% of this data. To aggregate and clean the 70% of data that streams in from third-party sources, most data management teams use manual methods that involve the EDC, statistical computing environments, email, and numerous spreadsheets.

These manual methods increase effort, costs, and potential risk. Taking data offline for cleaning delays its availability for periods that range from a few days to a few months, preventing more agile responses if safety or quality issues arise.

Faster, automated approaches are needed to speed the availability of clean, consistent clinical data. An emerging category of solutions, clinical data workbenches (also referred to as clinical data platforms, hubs, or data aggregation and management systems), addresses these challenges.

What are clinical data workbenches?

Clinical data workbenches provide a single source of truth for all forms of clinical data, making it faster and easier for cross-functional teams to manage the diverse patient data in today's clinical trials. Workbenches address challenges posed by the growing volumes of data from external sources like labs, wearable devices, and ePRO solutions.
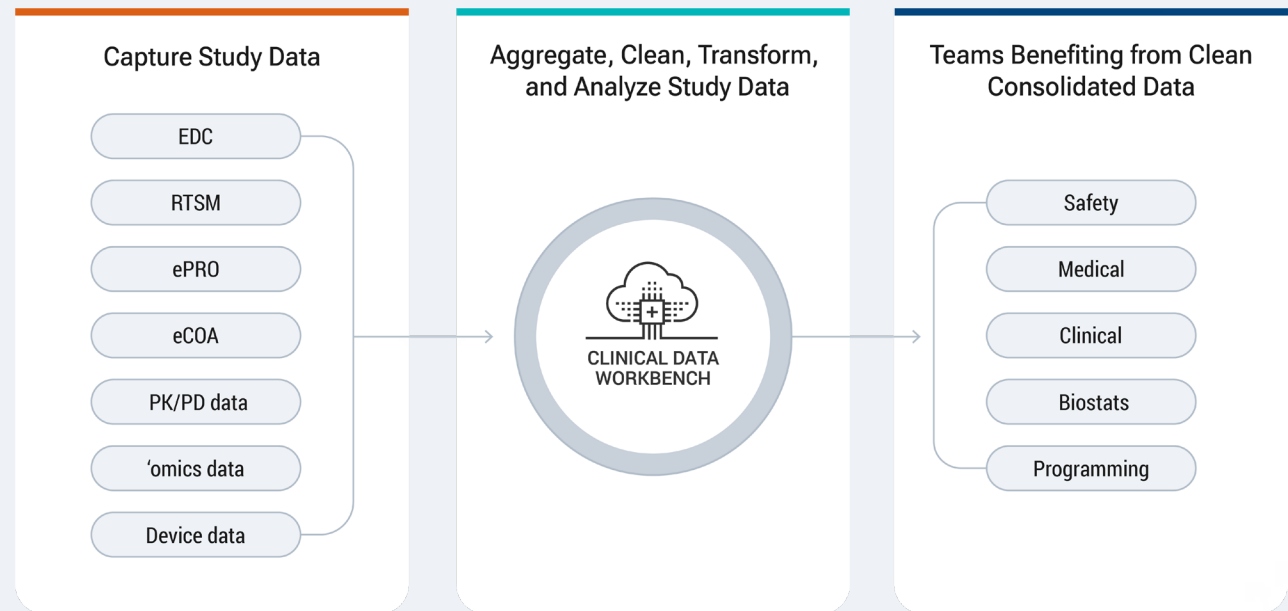
They centralize and harmonize the data, and use automation to reduce reliance on manual processes for transformation and cleaning. Workbenches also enable better use of analytics, ensure data integrity and quality, and improve collaboration among different stakeholder groups, thus facilitating increased trial agility and speed-to-market.

Providing a single, central location for all trial data **[Figure 1]** improves accessibility for adjacent and downstream processes such as safety surveillance or medical and clinical reviews. It also establishes a unified, harmonized foundation for clinical data to ensure the successful application of AI and other emerging technologies.

While the first workbench appeared on the market years ago, the first-generation systems were never widely adopted. Today, there are new offerings from a diverse range of providers, including Edetek, Medidata, Oracle, Saama, Signant Health, and Veeva. Each employs a different model and approach. Across the category, workbenches serve a wide range of functions, most of which deliver some but not all of the following: aggregation, review, query management, transformation, storage, visualization, and artificial intelligence.

FIGURE 1

Clinical Workbenches Support Diverse Data Sources and Consumers

Capture Study Data

- EDC
- RTSM
- ePRO
- eCOA
- PK/PD data
- 'omics data
- Device data

Aggregate, Clean, Transform, and Analyze Study Data

CLINICAL DATA WORKBENCH

Teams Benefiting from Clean Consolidated Data

- Safety
- Medical
- Clinical
- Biostats
- Programming

The technological advances in this latest generation of clinical data workbenches deliver a better return on investment than their predecessors.

Sponsors and CROs report seeing improvements from modern workbenches, including fewer manual processes, automated change detection, external patient data verification, and reduced cycle times for query management and database lock.

This guide offers a broad overview of data workbench technology, summarizing fundamentals for understanding basic system requirements, developing a business case for investment, and selecting the workbench best suited to your business needs.

Determining what you need from a workbench

The first step in evaluating clinical data workbenches is deciding what your organization needs and how it could benefit most from this application. The checklist below summarizes the basic business requirements to consider.

If the data workbench will be a net new system, your data review and cleaning processes should change to leverage its capabilities. Thus, when defining requirements, "ask for a car, not a faster horse." Instead of seeking to improve a legacy process, strive for greater advances by investigating what's possible with these new systems. But verify vendor claims to ensure your "faster horse" is not a "flying car."

System Requirements

This checklist groups prospective capabilities into the following business requirements:
Data Aggregation, Data Cleaning, Data Transformation, Data Analysis, and System Infrastructure.

Data Aggregation

☑ **DATA INGESTION**
Your data workbench should accept trial data from every source used in your trials and require little to no effort on your part. Be sure that it will ingest data at the frequency you need. Productized integrations between the workbench and preferred data providers add extra value because the providers themselves support them.

☑ **EXCEPTION HANDLING**
Upon data ingestion, the system should automatically detect issues within the data structure and in the data file itself. Ask the vendor what types of issues and errors can be automatically identified. Warnings should be made available to your data management team and vendor.

☑ **DATA MAPPING**
Ask vendors to describe their data mapping strategies. Utilizing a simple data model will reduce the number of data mappings that are required. Fewer mappings mean that fewer things that can go wrong and delay the start of data cleaning and review.

☑ **DATA STANDARDIZATION**
Standardization generally simplifies downstream activities, but the required data transformations come at a cost. Strike a balance between easing downstream activities and making data available quickly for cleaning and review.

Data Cleaning

☑ **LISTING CREATION**
Most users want to view data in a tabular form (i.e., as listings), and a workbench should produce tabular layouts quickly and easily. No-code methods are user-friendly and empower data managers, although sometimes complexity cannot be avoided. Thus, look for both code and no-code capabilities to create listings.

☑ **AUTOMATED CHECKS**
Automated checks save time and speed up the cleaning process by automatically identifying data discrepancies and raising queries about them each time new data is loaded. Look for code and no-code options for writing the scripts. Automated checks contribute significantly to a workbench's ROI, so the greater the capability, the better.

☑ **QUERY MANAGEMENT**
Workbenches centralize query management. You should be able to manage and action queries across multiple data sources. The ability to track existing queries when data is refreshed will prevent users from wasting time updating spreadsheets or re-reviewing data unnecessarily.

☑ **QUERY ACCESS**
When centralizing query management, you need mechanisms to support communications between systems and companies. Look for the ability to push and pull queries from external systems, including the EDC. External data providers should be able to log in to the workbench to view and resolve queries about their data.

Data Transformation

☑ **SDTM OR SDTM-LIKE OPTIONS**
While we don't recommend performing data cleaning on SDTM datasets, the cleaned data must be transformed into a standardized structure. Consider whether you want to create SDTM or SDTM-like domains within your clinical data workbench.

☑ **DATA EXPORTS, BOTH AD HOC AND SCHEDULED**
A workbench should be a conduit through which study data flows. Robust scheduling capabilities allow it to become part of an integrated process flow, so downstream users can have data exported automatically as often as needed. It should offer both push (for ad hoc and scheduled exports) and pull mechanisms (via an API).

Data Analytics

☑ **OPERATIONAL DASHBOARDS**

Dashboards provide visibility into the completeness and timeliness of cleaning activities. Their tracking and metrics enable leaders to actively manage the process, improve cycle times, and deliver better service to downstream data consumers.

☑ **DATA VISUALIZATION**

Visualizations make it easier to identify anomalies and trends. As with anomaly detection, you may wish to layer visualization capabilities on top of your aggregated and cleaned data. This capability may exist within your workbench or via data export capabilities that push the combined data to a dedicated visualization system.

☑ **AI/ML ENABLED DATA REVIEW**

Artificial Intelligence and Machine Learning capabilities can help identify data anomalies and atypical values at scale. Surfacing anomalous values helps reviewers focus their efforts. Reviewers should be able to drill down from auto-generated lists to the underlying data for further investigation and to initiate queries as needed.

☑ **CONSOLIDATED EXPORTS TO AI/ML SYSTEMS**

Organizations using AI/ML for clinical analyses can benefit from the clean, consolidated data produced by data workbenches. Ensuring that ML models for RBQM are trained against clean data improves the output quality. Ensure your workbench provides granular controls and scheduling capabilities for data exports.

☑ **PATIENT PROFILES**

Your workbench should provide a comprehensive and up-to-date status of data cleanliness for all data collected at the individual patient level. A patient-centric view allows data managers to prioritize time and resources and lock patients faster for interim analyses, giving medical and safety teams visibility and perspective for better decisions.

System Infrastructure

☑ **SCALABILITY**

When evaluating scalability, estimate the number of studies you will run over the next five to seven years and how much data they will contain. The workbench must be able to handle the highest volume of data expected from your largest studies, as these are often the most important, and have the most to gain from a workbench application.

☑ **PERFORMANCE TESTING**

Any workbench may perform well when demonstrated on a single study that involves small volumes of data, but how will it hold up in a mega trial? Ask vendors about their performance testing and for references from customers who've used the technology to handle similar volumes of data.

☑ **SYSTEM INTEGRATIONS**

Workbenches sit between data providers and data consumers; therefore, an open, well-documented API is imperative, and productized integrations are highly valuable. Engage your IT team to evaluate how easy or difficult it will be to integrate this workbench with other systems.

☑ **SERVICES AND TRAINING**

Are you buying an off-the-shelf product or a services-based solution? Once the workbench has been implemented, will you need ongoing vendor services? If so, why? What level of service will be required for each study? How much training would your teams need to become self-sufficient in configuring the system for each new study?

☑ **SLAs AND SUPPORT**

Ask about the Service Level Agreements (SLAs) related to system up-time and help desk responses. If your studies never sleep, your workbench system and its support team shouldn't either.

☑ **DATA SECURITY AND COMPLIANCE NEEDS**

Having aggregated all of your study data in one place, you need to be confident that it is secure. Make sure any prospective workbench vendor passes a rigorous security audit.

- Do they have systematic approaches to security architecture and follow best security practices from an organizational and personnel perspective?

- Do they have clear network communications and encryption practices and perform regular vulnerability and penetration testing?

- Have they established procedural information privacy controls?

- Are they open and clear about how they'll look after your data?

Maintain a healthy dose of skepticism when evaluating options

Some solutions in this space rely heavily on services-developed customizations to deliver functionality. Customizations can meet highly specific requirements and provide cutting-edge capabilities, but they also result in greater fragility. That fragility increases the ongoing maintenance costs for new trials and software upgrades.

Approach workbench evaluation with a healthy dose of skepticism and request external validation and proof of any messaging claims. Place extra value on referrals and testimonials from current satisfied customers, along with the ROI they have achieved.
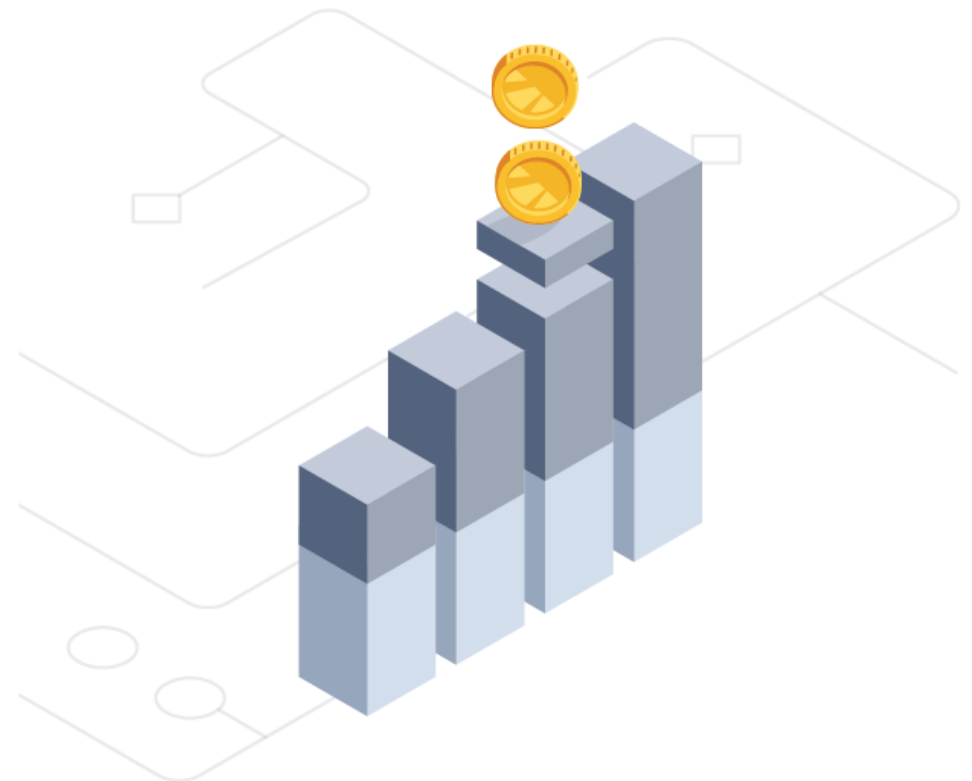
Ensure that vendors are transparent about the following:

• What their products can and cannot do today, out of the box.

• What can be done with customization.

• What they plan for future releases.

• What features have been delivered over the past 12-18 months.

Since services teams can add virtually any capability, services-heavy offerings can deliver false positives to RFP questions asking: "Can your system do…?" Ask vendors to specify which capabilities are part of the "out-of-the- box" offering and which are achieved through customization.

What is the pricing, and what are the commercial terms?

As you'd do when comparing electric and gasoline-powered vehicles, evaluate the Total Cost of Ownership (TCO) rather than just upfront licensing costs. A workbench's TCO should include licensing, implementation and training costs, projected system integration costs, maintenance fees, renewal costs, and all likely service costs over five years.

Making a business case

Gaining approval for a new business system takes work. To increase the likelihood of approval, articulate existing challenges and their impact, along with the anticipated value of your proposed solution.

Focus your business case on big-ticket items for the key decision makers: for programming, it's the time spent to generate listings; for data management, it's the manual validation checks and query management; and for safety and medical teams, it is the impact of delayed data.

A good business case should include the following five items:

1 PROBLEM STATEMENT

Define the specific business problems and pain points the new system will address. List the challenges and delays inherent in manual processes and bring them to life with quantification and anecdotal examples. Show the limitations and impact of the existing processes.

2 PROPOSED SOLUTION

Outline the proposed solution and its key features. Explain how these features will address the identified problems and deliver tangible benefits to the organization. Describe the system's core capabilities and alignment with the organization's strategic goals.

3 QUANTIFIABLE BENEFITS

Quantify the expected benefits of the new solution in terms of cost and time savings, productivity improvements, and quality-of-service measures. For example, how quickly can medical and safety teams access reliable data? Use quantifiable gains, such as fewer hours spent on manual validation checks, to demonstrate the system's potential return on investment (ROI).

4 IMPLEMENTATION REQUIREMENTS

Define the key steps to deploy the system, train users, and deliver the change management needed for a smooth transition. At this point, a detailed implementation plan isn't realistic. For the business case, define a realistic plan that provides a reasonable estimate of the resources required. Get input from IT on past costs based on the expected effort required.

5 SUSTAINABILITY PLAN

Describe the ongoing support and maintenance needed for the new system. Outline the costs associated with ongoing support, upgrades, and security measures. If incremental configuration or service costs are necessary for each new study, ensure those are captured and included. A realistic sustainability plan contributes to long-term system viability and a credible business case for making the investment.

Take action

Clinical data workbenches are gaining the attention of clinical data management leaders. They provide an attractive solution to challenges that have existed for many years and that are only getting worse. If you don't already have a workbench, now is the time to start planning for one.

Clinical data workbenches offer an alternative to legacy SAS-based manual aggregation, transformation, and cleaning to help ensure clinical data quality, regulatory compliance, and patient safety. The latest generation of workbenches is already demonstrating ROI by saving time, improving efficiency, and enriching collaboration between clinical trial stakeholders.

Delivering value today and tomorrow

As trial designs become more complex and the volume and variety of sources for patient data increase, workbenches are emerging with the data science tools needed to clean, transform, and analyze clinical data faster and with less effort. Study teams gain quicker access to a stream of clean and consistent data, helping them be agile and make better decisions, contributing to patient safety and trial efficiency.

In addition to providing immediate efficiency gains, a workbench provides a foundation for adopting new technologies that support adjacent business processes. Today, it is difficult to layer AI/ML and other emerging technologies on a siloed, manual infrastructure for clinical data. A modern foundation for clinical data readies the organization for future advances in data management and beyond.

This **eBook** provides more information on the current clinical data workbench landscape.

Veeva

About us

Veeva is the global leader in cloud software for the life sciences industry. Committed to innovation, product excellence, and customer success, Veeva serves more than 1,100 customers, ranging from the world's largest biopharma companies to emerging biotechs. As a **Public Benefit Corporation**, Veeva is committed to balancing the interests of all stakeholders, including customers, employees, shareholders, and the industries it serves.

For more information, visit www.veeva.com