Clinical Data Workbenches: An Evolving Product Category Speeds Access to Quality Data

Veeva

Growing volumes of third-party patient data have driven development of the clinical data workbench, designed to ensure data quality, enable better use of analytics, and reduce reliance on manual aggregation, cleaning, and transformation.

Any clinical trial is only as valid as the data that supports it. Today, establishing that validity is more difficult, since 70% of the patient data used in most trials comes from a diverse range of external sources, including labs, wearable devices, and electronic patient-reported outcomes (ePRO) solutions. It does not fit neatly into a CRF and cannot be managed in an EDC, challenging data managers to find and resolve inconsistencies quickly or efficiently.

So far, companies have responded by deploying large teams of data checkers to aggregate, transform, and clean this data. Global teams work around the clock, using manual methods, statistical computing environments, numerous spreadsheets, and countless email exchanges to accomplish these tasks.

But even these costly efforts are unlikely to prevent trial delays or to keep up with increasing volumes of external patient data and more complex trial designs. Genomics trials today, for example, can link terabytes of data to primary endpoints, and the industry needs better ways to clean and access all that data.

Even worse, manual tasks are often performed inconsistently, and poorly documented. This can raise concerns about data quality and invite regulator scrutiny.

To speed stakeholder access to trusted data from all sources, clinical data scientists are focusing on key areas. Upstream, they are paying closer attention to data ingestion, cleaning, transformation; downstream, they are concentrating on clinical analytics (e.g. visualization) and operational analytics (e.g., RBQM and dashboards).

These upstream and downstream capabilities are part of an emerging product category that:

- Centralizes complete and concurrent patient data from all sources on one platform
- Permits closer oversight—not just of data integrity, but of data quality—to enable better use of analytics and visualization of real-time trends
- Empowers clinical data teams to manage outcomes using technology, instead of using manual processes to manage tasks
- Improves collaboration by enabling consistent workflows across stakeholder groups

Creating a pipeline of trusted data across varied sources and functions

Within the past two years, the shift from clinical data management to clinical data science has emphasized the need for prompt decisions with quality data. The requirements for getting this data have shaped an evolving, and often confusing, category of data management solutions. Typically referred to as clinical data workbenches, and sometimes as clinical data hubs or platforms, they are designed to reduce the cost as well as the potential error and noncompliance of manual data aggregation, reconciliation, transformation, and analytics.

By enabling faster access to a pipeline of quality study data across functions (Figure 1), data workbenches are already improving cross-functional collaboration and increasing trial agility by feeding clean and harmonized data to visualization and RBQM tools. Some workbenches already include visualization and analytics capabilities, a trend that is likely to continue in the future as designs improve.

Today, when external patient data predominates, reliance on the wrong tool creates risk and vulnerability. For example, a statistical computing environment may aggregate and transform data well, but won't enable data cleaning; a clinical data repository may aggregate but cannot transform, while a visualization application won't handle aggregation effectively.

FIGURE 1

Desired Collaboration Using a Clinical Data Workbench

Spreadsheets can't make tools fit-for-purpose

Typically, data management teams use EDC to collect data, leveraging the query workflow to manage data discrepancies. But, in order to identify discrepancies across EDC and other data sources, they've had to extract data from the EDC, put it into a data repository, read the data from the repository into a statistical computing environment, and then program specialized reports and email them to reviewers as spreadsheets.

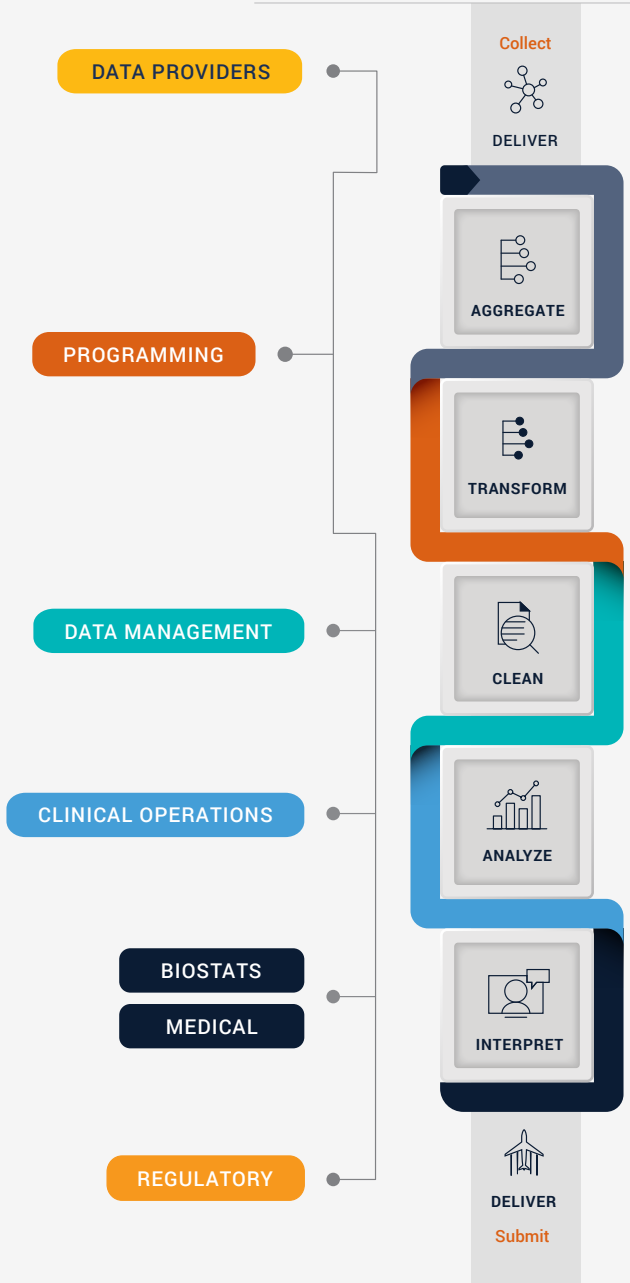
Reviewers, in turn, have had to find each discrepancy, log back into the EDC, and type messages to research site staff or data providers to resolve.

As one expert puts it, "If you pick one data point in a study you've likely cleaned it 17 times on average throughout the course of that study. It's not the right way to use people's time and delays patient access to new drugs." A multi-source patient data environment demands more robust and streamlined approaches that eliminate redundant processes and provide a holistic view of the data.

Clinical data science requires deeper collaboration between different stakeholder groups, such as clinical operations, clinical data, biostatistics, safety, and medical experts. A clinical data workbench will allow each group to bring its own perspective and work productively, and in tandem, on the same concurrent data (Figure 2).

FIGURE 2

The Goal: Clean, Consistent Data, Accessible Across Functions



AGGREGATION

- Ingestion by source or data provider, including metadata
- Data harmonization with common definitions

TRANSFORMATION

- SQL/CQL programming
- SDTM transformation

CLEANING

- Reconciliation
- Query management

OPERATIONAL ANALYTICS

- RBQM
- Operational dashboards

CLINICAL ANALYTICS

- Visualization
- Patient profiles
- RB-CDM

Creating value and ROI

Today's data workbenches are already proving their worth and generating ROI, especially when attached to practical use cases. For some companies, being able to harmonize data in one place has sped ingestion and aggregation.

In transformation, workbenches have reduced manual effort and eliminated the need for custom programming; in review, they're eliminating the use of trackers, and automating nearly one third of data verification steps, reducing the time required for query management by more than 60 hours per week. They've also helped companies adopt visualization and analytics technologies successfully, and to prep data for AI/ML use cases.

This eBook looks at the role that clinical data workbenches will play in ensuring access to quality data as volumes of diverse, multisource patient data continue to increase.

Examples of ROI from clinical data workbenches

- Allow queries to be closed without human intervention
- Eliminate the need for spreadsheet trackers
- Eliminate listing programming requirements
- Automate 30% of the steps required to verify external patient data

Surveying the technology landscape

As shown in Figure 3, many companies are using a separate tool for each of the elements in the data management landscape. This will centralize over the coming years.

Role-based system capabilities and permissions enable a data workbench to support multiple steps and users in the data lifecycle. Within this environment, individuals can access the latest data for review, reconciliation and visualization.

The next step in the process is utilizing the data for interpretation and decision making by clinical data experts and business leaders. These activities evaluate real-time trends and are distinct from the clinical reports used for regulatory submissions.

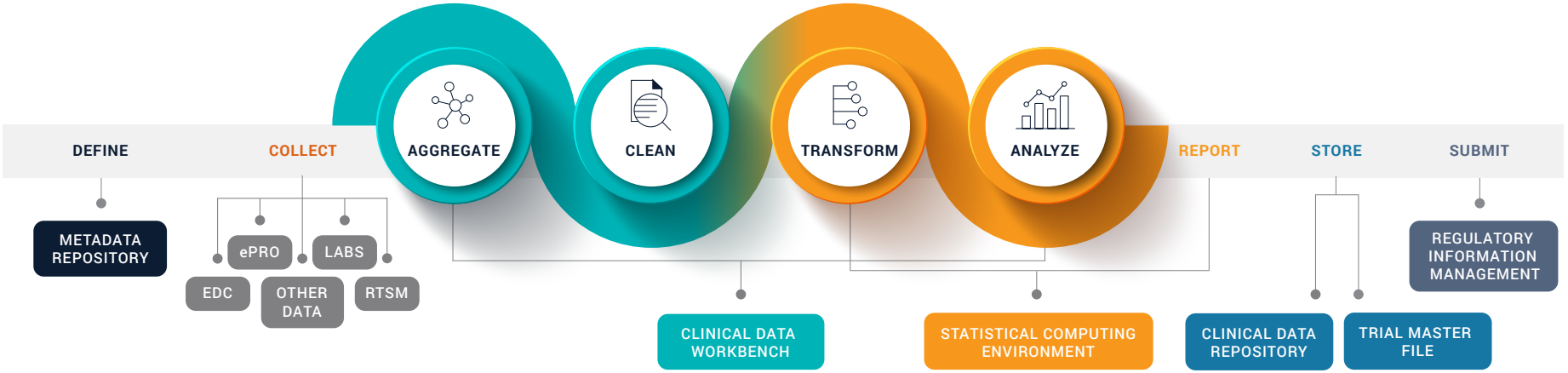
Increasingly, risk-based processes such as RBQM and RB-CDM, as well as predictive and prescriptive approaches, are also being used, as machine learning and artificial intelligence techniques are deployed into clinical data workflows to inform clinical and operational decisions. These analytics allow in-stream decisions to be made to answer such questions as: 'Does the site need more training to address high issue rates? Should this site be used in a future clinical trial? Is a new safety signal emerging?'

Clinical reports, meanwhile, examine data to support marketing claims and applications for regulatory approval. They are designed to answer such questions as whether the drug benefit/risk ratio is acceptable for a given population, or whether the target drug demonstrates efficacy when compared with a placebo or comparator.

After analysis, clinical reports are included in the eCTD for regulatory submissions, after which all the trial data and documentation is stored and archived, typically on a separate technology platform from the one used for aggregation and cleaning.

FIGURE 3

The Clinical Data Product Landscape



AGGREGATE

Ingest, consolidate, and harmonize data across sources. Simplify data acquisition methods across all data providers / data sources.

CLEAN

Centralize data cleaning and review in a single solution for all study data. Securely manage access and blinding. Automate data issue detection and cleaning.

TRANSFORM

Deliver data in an interoperable package to any downstream system. Speed up data availability / reduce lag for stakeholders.

ANALYZE / REPORT

Achieve insights and reach decisions faster. Use reports and dashboards to evaluate real-time trends in data collection, site, and vendor activity, as well as data management and clinical results. Separately, analyze clinical data to support regulatory applications and marketing claims.

DEFINE

Create re-usable data collection, transformation, and reporting definitions. Relate study data elements to each other and to defined standards. Reduce re-work, increase re-use.

COLLECT

Collect study data as per protocol requirements. Create easy and enjoyable site and patient experience.

STORE

Archive and store data and audit artifacts. Maintain compliance with regulatory data integrity requirements.

SUBMIT

Reach time-to-market faster. Provide treatments to patients. Reduce study costs.

Walk before you run: prioritizing operations

When approaching new technology decisions and process improvements, it is tempting to tackle too much at once. The risks to project success and adoption compound when bringing net-new technology to your organization.

Therefore, during evaluation, prioritize business objectives and consider the natural order of operations.

The clinical data lifecycle builds on each step before it. The effort required in downstream steps depends on the quality delivered from upstream. Only after data has been aggregated, cleaned, and transformed can one successfully implement visualizations, risk-based tools, or AI/ML approaches.

Addressing the challenges of multisource patient data accelerates the need for a clean data foundation. Applying AI/ML algorithms to bad data can lead to false conclusions and added risk, so focus first on getting quality data with speed and efficiency.

In such an environment, it makes sense to optimize processes and prioritize investments from back to front in the data flow.

Defining the future state

For sponsors and CROs, the availability of clinical data workbenches offers a major advance in operating efficiency. Traditionally, they've had to manage clinical data in dribs and drabs, like a relay race in which each person transfers small increments of data to the next teammate. It requires painstaking manual effort, with the hope that nobody drops anything.

Clinical data workbenches change that picture and bring data managers into clinical data science, empowering them to move beyond using manual methods to handle tasks, to using technology to help manage outcomes. This aligns well with SCDM's vision of clinical data management moving to clinical data science, and will allow teams to analyze data sooner than they can now, and spot problems earlier in a trial.

Effectively adopting a clinical data workbench speeds time to submission by delivering a pipeline of quality data, flowing from end-to-end and accessible for all clinical trial stakeholders.

Learn more about Veeva's clinical data workbench solution.

REFERENCE

Alani, L., Bain, D. et al., "Functional Integration Map," Association for Clinical Data Management, 2023.