ORACLE
HEALTH SCIENCES

Accelerating Data Cleaning, Integration and Analysis from Source to Submission

How Oracle Health Sciences Data Management Workbench delivers true end-to-end clinical data management

ORACLE WHITE PAPER | SEPTEMBER 2015

ORACLE

Disclaimer

The following is intended to outline our general product direction. It is intended for information purposes only, and may not be incorporated into any contract. It is not a commitment to deliver any material, code, or functionality, and should not be relied upon in making purchasing decisions. The development, release, and timing of any features or functionality described for Oracle's products remains at the sole discretion of Oracle.

Table of Contents

.

Executive Summary

Clinical organizations are under increasing pressure to execute clinical trials faster with higher quality. Subject data originates from multiple sources; CRFs collect data on patient visits. Core labs return blood test results. Implantable medical devices such as insulin pumps collect continuous patient data. Other examples, adding to the complexity, include electrocardiograms, gene sequencing, and wireless technology. All this data needs to be integrated, cleaned and transformed from raw data to analysis datasets. This data management across multiple sources is on the critical path to successful trial execution, and several challenges undermine its long-term sustainability.

First, most organizations face increasing data volumes and number of IT systems, as a result of incorporating more external data sources (e.g., devices) in their clinical trials. Second, cleaning and transforming data across multiple sources is a complex activity which involves multiple roles and interdependent tasks, whose coordination is regulated by standard operating procedures (SOPs) which often involve manual processes hard to follow and track. Third, adherence to multiple standards is becoming a priority, as organizations seek more insights from pooling multiple studies, as well as compliance to newly mandated requirements from the regulatory agencies. Overall, there is a need for moving from a world characterized by fragmented knowledge and processes driven by spreadsheets, to a world where one central IT system controls and enables automation for clinical data management across multiple sources.

In this white paper, we describe in detail the key challenges of clinical data management across multiple sources, and explain how organizations are using Oracle Health Sciences Data Management Workbench (DMW) to overcome these challenges.

Dealing With Multiple Data Sources

It is currently estimated that the average time to develop a drug is more than 10 years, with an average cost in the last decade of over $2 billion per drug.[1] Most of this time and money goes into running the clinical trials needed to test the new treatment and produce supporting evidence of the expected benefits. Therefore, life science organizations are trying to get more out of their clinical trials, by increasing the number of data points they collect, to get a better picture of what's happening with the patient during treatment administration.

For example, some insulin providers started using continuous glucose monitoring in their trials, significantly increasing the amount of data available on each patient. This increase in data volumes and external data sources can increase risk, such as:

- Delays in creating regulatory submissions resulting in material commercial impact
- Unsustainable clinical trial throughput because of increasing clinical trial activity
- Increased risk of regulatory non-compliance which may result in rejected or delayed submissions

Workarounds that clinical organizations often adopt to attempt to address these demands include:

- Increasing computing power
- Building custom integrations
- Increasing workforce
- Reorganizing IT departments

Although all potentially valid, these efforts are destined to fail if they don't address the prime cause, which is that data management is a complex activity which involves multiple roles and interdependent tasks. Increasing computing power is a necessity, but even the most high-performance system can't enable people to handle extra work when the bottleneck is in manual processes regulated by SOPs but not supported by automation. Custom integrations are a two-faceted sword – they "do the job" but they are costly to maintain. Increasing workforce may not be possible in all organizations, and even when possible it may add to the complexity of the communication lines. Reorganization can help simplify the process but it doesn't shorten the time of tasks that are on the critical path (e.g., creation of SDTM datasets).

The overall objective of executing trials faster with high quality can then be reframed as: How can clinical organizations handle data management across multiple data sources? How to shorten time from collection of raw data to clean submission and analysis datasets? How to meet regulatory requirements, like secured access, traceability and blinding? How to handle trials with larger data volumes, higher number of data sources or more complex design (e.g., adaptive trials)?

End-To-End Clinical Data Management Process

When dealing with multiple data sources, it is desirable to consolidate these data sources into one clinical data repository to provide a single and consistent source of truth for data, programs, outputs and reports. During study execution, the raw datasets may be transformed into a variety of formats, such as CDISC (SDTM, ADaM), analysis-ready datasets, or standards internal to the clinical organization. Ultimately, these datasets are analyzed for reporting (Tables, Figures and Listings), to pool data from different studies and identify trends, or to support other business operations (e.g., risk-based monitoring). Consolidating data into a clinical data repository enables data management activities as well as a broad cross-section of business questions to be answered (see Figure 1).

1 Tufts Center for the Study of Drug Development, November 18, 2014: http://csdd.tufts.edu/news/complete_story/pr_tufts_csdd_2014_cost_study

When dealing with multiple data sources, it is desirable to consolidate these data sources into one clinical data repository to provide a single and consistent source of truth for data, programs, outputs and reports.

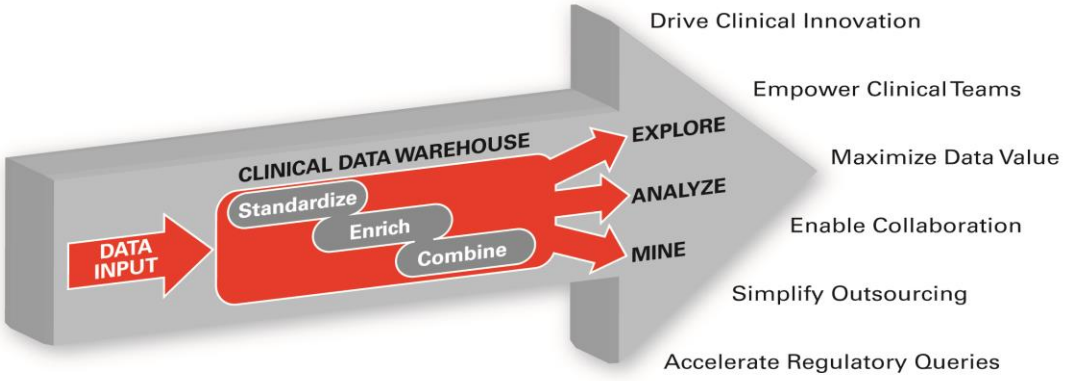

Figure 1. Overview of end-to-end clinical data management activities and expected value.

Across the clinical data chain there is the need to integrate data at different stages in the lifecycle, and enable access to a wide group of stakeholders including Data Management, Medical Review, Clinical Operations and Biostatistics – both internally and externally (see Figure 2). Timely integration and aggregation of data is essential to produce business-critical analyses to drive operational management of the trial, as well as supporting efficacy and safety conclusions.

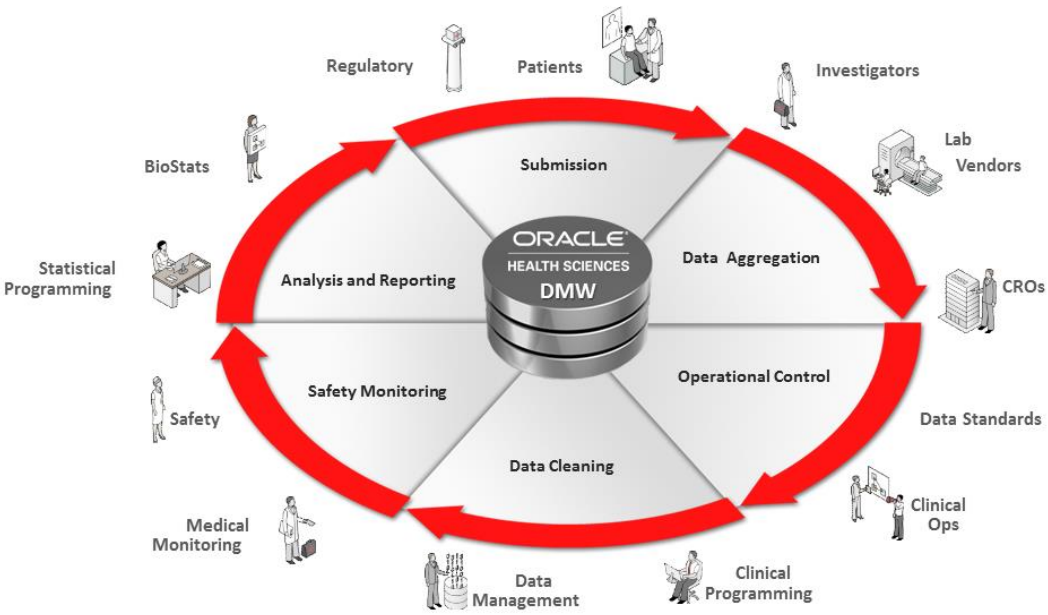

Figure 2. Collaboration of different stakeholders in a clinical trial.

Oracle Health Sciences DMW addresses this process with data cleaning and transformations from different data sources. It provides one solution for data management and integration of different data sources, and supports business needs around consolidation and distribution of clinical and operational data.

As standard functionality, DMW integrates data from several sources and formats, such as Oracle Health Sciences InForm EDC, SAS datasets, and text files. Once source data is acquired, DMW provides a single source of truth for consolidated data, to support data cleaning, transformations, and downstream data analysis and reporting.

Further, Oracle Health Sciences DMW has bidirectional integration with Oracle Health Sciences InForm EDC platform. Data discrepancies can be raised in DMW and sent back to their originating system (e.g., InForm, third-party data providers such as labs), thus providing a central place for discrepancy management. The integration of Oracle Health Sciences InForm with DMW allows the data capture of core EDC-based data to be rapidly consolidated, reconciled and cleaned with non-EDC data to confirm robust curation of all data across the study, independent of source.

Compare this with what most clinical organizations do today: copying data locally and running programs to generate spreadsheets, which are then manually reviewed by data managers or medical reviewers, and where the resulting queries are manually transcribed into the EDC. With DMW, when your study team identifies a potential data collection issue, they can seamlessly raise discrepancies within DMW, assign categories, and set actions to push discrepancies out to the appropriate individuals—or to InForm—for rapid resolution (see Figure 3).

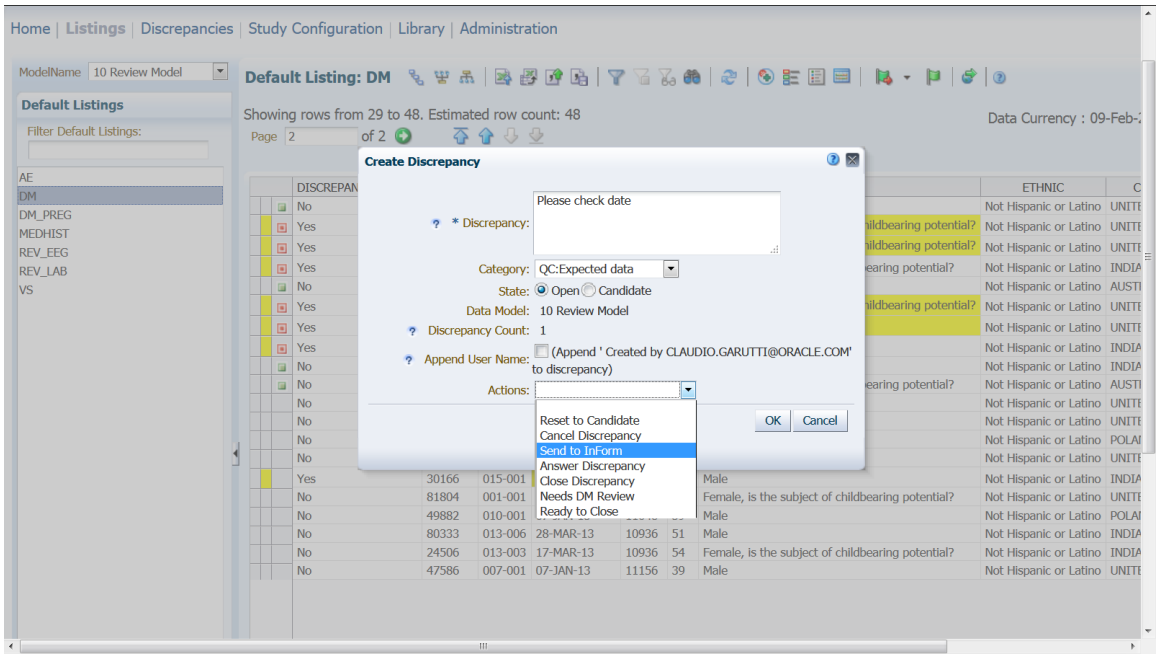

Figure 3. Oracle Health Sciences Data Management Workbench (DMW) enables the cleaning of multiple data sources faster and with higher quality through its productized integration with Oracle Health Sciences InForm EDC.

The integration of Oracle Health Sciences InForm with DMW allows the data capture of core EDC-based data to be rapidly consolidated, reconciled and cleaned with non-EDC data to confirm robust curation of all data across the study, independent of source.

For data managers that require ad hoc reports, Oracle Health Sciences DMW lets you quickly create and share custom listings to identify more complex data anomalies. The custom listings sit on top of the single source of truth, avoiding data duplication and orphan spreadsheets that travel from Inbox to Inbox, and therefore avoiding confusion on what has been processed and what is still pending resolution. This helps the study team avoid wasting time chasing internal inconsistencies, helping to increase both productivity and quality.

In addition, DMW's comprehensive security features preserve any data "blinding" rules across integrated platforms—ensuring trial staff and data managers remain fully compliant during randomized clinical trials and data cleaning operations. The granularity of the blinding is down to the single data point. This allows even the most complex protocol to be implemented in a real end-to-end compliant system (see Figure 4).

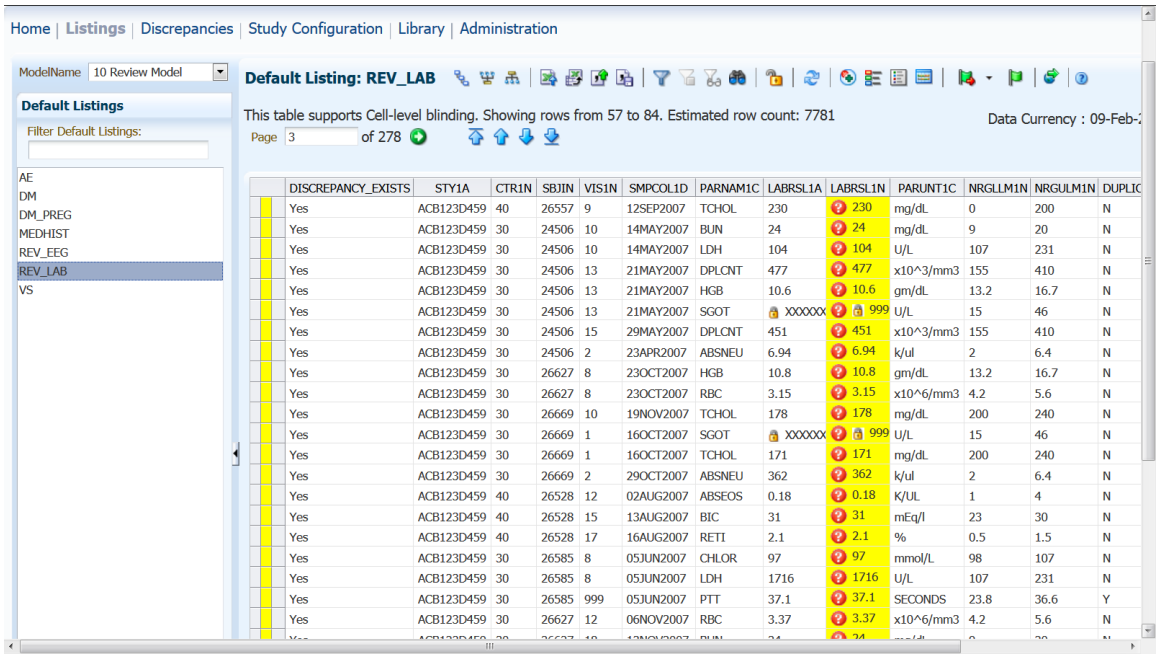

Figure 4. Granular blinding in DMW – down to the single data point - makes it easier to conduct clinical data management in compliance with the protocol.

Utilizing a metadata-driven transformation engine, DMW accelerates clinical data management by allowing data managers to merge, transpose, and stack together data from multiple sources into a standardized model–enabling cross-source review and analysis – resulting in higher-quality data and faster database locks. Trial efficiency can be increased even more by directly copying mapping rules used in previous studies, and increase standardization across your clinical trials for drastically reduced study setup cycle times.

Oracle Health Sciences DMW accelerates clinical data management by allowing data managers to merge, transpose, and stack together data from multiple sources into a standardized model—enabling cross-source review and analysis — resulting in higher-quality data and faster database locks.

When working on data validation, you can trace the lineage of your data at any given point within DMW to identify how the data was mapped, and follow data pathways back to Oracle Health Sciences InForm EDC platform for fully transparent auditing and traceability (See Figure 5).

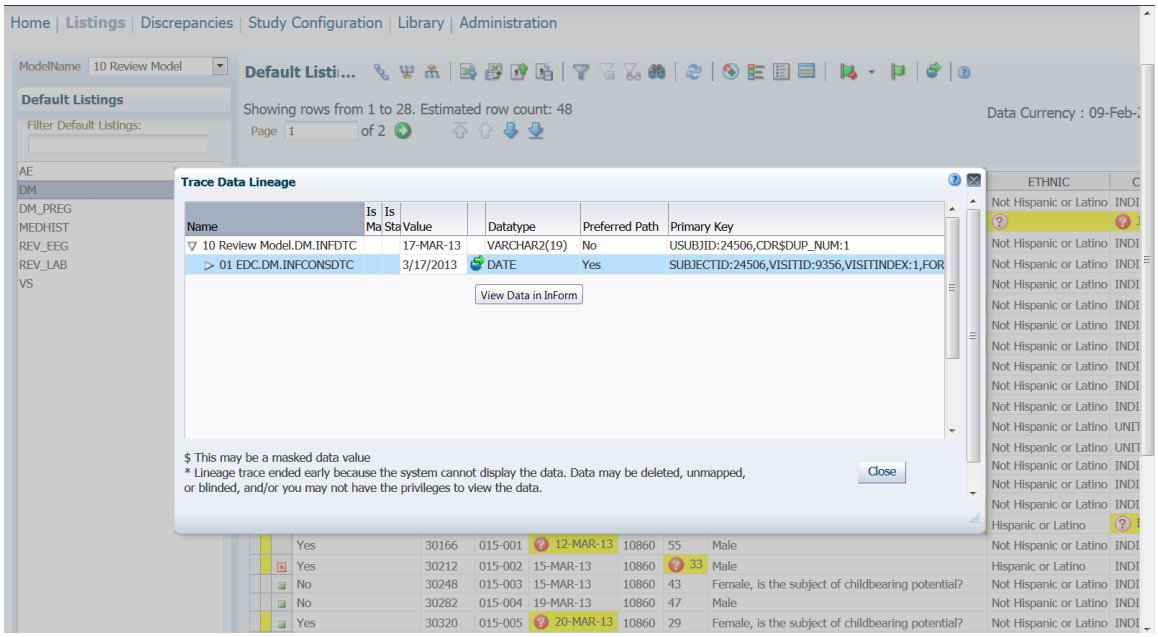

Figure 5. End-to-end traceability makes validation from source to derived datasets faster and easier.

Medical coding is also embedded in DMW, which is integrated with Oracle Thesaurus Management System (TMS).[2] Users can store dictionaries and terminologies in TMS (e.g., MedDRA, WHO-DD) and enable automatic coding in DMW. For example, adverse events in domain “AE” can be coded to their High Level Term and Low Level Term using a MedDRA dictionary stored in TMS. Adverse Events that don’t have a match in the selected dictionary can be automatically queried, streamlining data cleaning. Multiple MedDRA dictionaries can be stored in TMS and the same study can be easily recoded by associating it to a different MedDRA dictionary.

2 Oracle Thesaurus Management System website: http://www.oracle.com/us/products/applications/health-sciences/e-clinical/thesaurus-management/index.html

Productivity and Traceability – Everyone Wins From Process Optimization

Clinical organizations rely on numerous systems. Systems developed to meet discrete business processes often struggle to evolve and adapt with the dynamics of changing needs. The result is that such changes often drive a plethora of inefficient manual steps to fill in the gaps. While this approach may work for many years, it is often reliant on one core component, 'the worker'. SOPs may aim to enforce a process; however the team member will often execute a manual process in accordance with their own interpretation, ultimately opening up the organization to unpredictability, and subsequent risk. The phenomenon of 'swivel chair integration' is also common where someone uses system A to perform one task, and then manually transcribes the result into system B, by 'swiveling their chair' to the other system. Inherently, minimizing the number of systems within a business lifecycle – and maximizing re-use and optimization – is critical to both productivity and ensuring traceability.

Another key area for re-use is typically found within the area of clinical programming. Clinical organizations frequently rely on a handful of knowledgeable individuals who know their trials inside out. As these key employees change roles or move out from their organization, new clinical programmers (or biostatisticians) step in and hope to find neat documentation and linear code easy to digest. The truth is often the opposite. In one such situation, as a study progressed, the study team got more demanding in their requests for ad hoc reporting and data extracts, and the programs got more lengthy and convoluted, making its reuse more and more complex. One manager compared the task of reusing someone else's code as "looking at a black box with a flashlight". The lack of reusability is a key issue of current data management systems, and significant effort is spent in reproducing edit checks, mappings and reports that are similar across trials.

In Oracle Health Sciences DMW, study components are not seen as code, but as objects. Objects include data models, transformations, validation checks and custom listings. Once an object is created for a study, one can easily create a (linked) copy for another study. The link means that if the original object changes, than the user can decide to propagate the change to the copied object. Unlike code, object execution is independent of its location. For example, a user can implement a "join" transformation in study #1 that maps domains "DM" and "VS" into "DMVS". The same transformation can be easily copied into study #2, without the need to edit any code. DMW will compare the metadata of study #1 and #2, and suggest the mappings from study #1 which could be applied to #2. The same ease of reusability also applies for other objects like data models, validation checks and custom listings.

What about documentation? In DMW, implementation is self-documenting within the system. For example, the "join" transformation between "DM" and "VS" into "DMVS" will automatically generate documentation which tells us what is joined, on which variables, whether those are inner or outer joins, and any other information that uniquely identifies that transformation. If the study designer changes the transformation (e.g., merging on different columns), the documentation changes too.

In DMW, implementation is self-documenting within the system. For example, the "join" transformation between "DM" and "VS" into "DMVS" will automatically generate documentation which tells us what is joined, on which variables, whether those are inner or outer joins, and any other information that uniquely identifies that transformation.

Study component reusability and automatic documentation together enable clinical organizations to have greater clarity on what has been done to get from source (e.g., EDC, labs data) to target (e.g., SDTM) – in essence, to turn on the light in the black box. Ultimately, re-usable objects accelerates study setup, and combined with automation, reduces manual processes, and increases traceability.

End-To-End Standards for Regulatory Submission

Recent FDA Binding Guidance[3] has set the clock ticking to automated, standardized submissions, requiring traceability from collection to submission. DMW can assist in answering regulatory queries on the state of the data at any point in history, thanks to the traceability of its transformations and to its tight integration with InForm EDC.

Regulations require tight adherence to a specific standard within a submission, but specific versions have windows of support. Successful companies will support multiple standards across filings. DMW can speed up the changing of a version for pooling and submission, and enable recoding of medical terminology at the touch of a button.

With standardization and re-use comes automation: enabling the automation of transformations to and from review and submission models. DMW allows a company to realize their investment in electronic data capture, and automate the downstream processes to submission.

Change control, enhanced metadata management and end-to-end traceability in DMW enable:

- Faster, more accurate rsponses to regulatory queries
- Improved clarity of the chain of ownership to downstream users
- Powerful pooling capability
- Ability to make submissions in any supported version

Case Study: Top 20 Sponsor Embraces Integrated, End-to-End Clinical Data Management

A global biopharmaceutical organization wanted to re-tool its data collection, data cleaning, transformation, analyses and reporting structure and processes. Their data flow involved a mix of linear and manual processes, with business processes designed on a “per study” level, and multiple workarounds to address the problem of data not available to key stakeholders/users within appropriate timelines.

Its data management team was relying on more than one hundred ad hoc reports that needed to be created for every study due to lack of adherence to standards. Its SDTM mapping was done before entering their clinical data repository, meaning that if data was needed before SDTM was available, workarounds or extraordinary efforts were applied to get data earlier. From SDTM, additional manipulation of the data was needed to get to ADaM, making traceability difficult.

Having previously standardized on the Oracle Health Sciences InForm EDC platform, the company adopted Data Management Workbench to achieve its objectives for true and seamless clinical data management, including a) library of global standards, b) implementation of parallel business processes, c) enhanced traceability from source to SDTM and ADaM, meeting FDA requirements, and d) moving SDTM mapping after data review, therefore removing it from the critical path.

Their challenges were related to those of the industry: growing data volumes coming from multiple sources, complex data management process, and mapping to standards. Oracle Health Sciences Data Management Workbench provided the solution they were looking for, including:

- a highly scalable and secure clinical data warehouse that delivers peak performance
- productized integration with their InForm EDC system
- built-in data conformance and traceability, and robust and flexible blinding features
- a metadata-driven mapping engine which enabled them to move the SDTM mapping out of the critical path, reducing trial lead time

3 US Food and Drug Administration. Providing Regulatory Submissions In Electronic Format — Standardized Study Data: http://www.fda.gov/downloads/Drugs/Guidances/UCM292334.pdf

Use Case: Data Management of Adaptive Trials

Adaptive trials are clinical studies where one or more parameters (e.g., sample size) can change as the study progresses, in a pre-specified manner, without compromising its scientific validity. Across the industry, simple adaptive designs are used on approximately 20% of clinical trials.[4] Adaptive trial designs may save substantial time and resources by early study terminations due to futility; sample size re-estimation; and reduction of the number of protocol amendments. One of the main barriers to wider adoption of adaptive trials is operational concern, in particular delays and disruptions in trial execution.

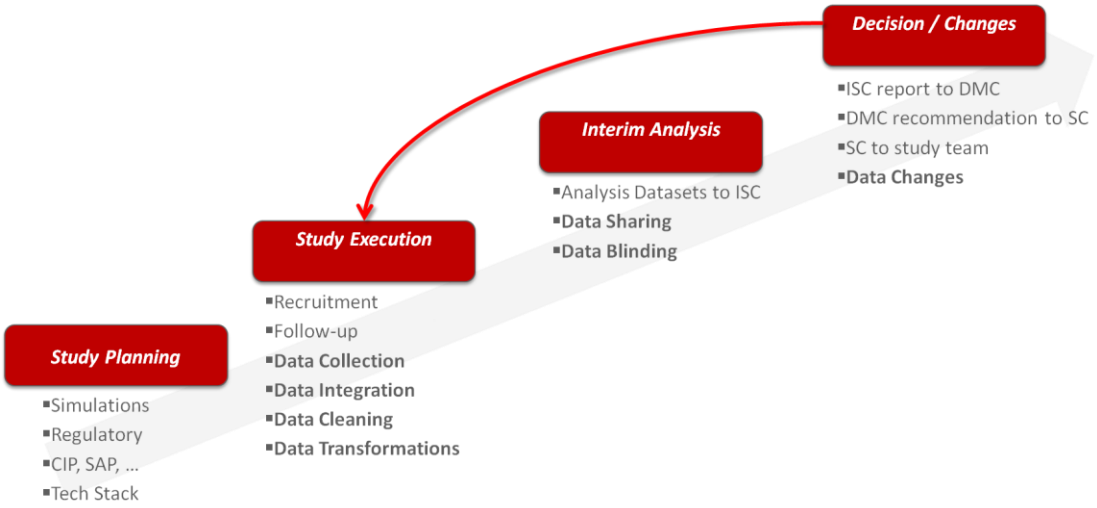

Figure 6. Example of study flow in adaptive trials. Boldface indicates data challenges optimized by Oracle Health Sciences Data Management Workbench.

Figure 6 illustrates adaptive study conduct. Study planning and execution are common to traditional studies; the new elements are the interim analysis and the decision that follows the interim analysis. These two steps can be repeated multiple times, as illustrated by the backwards arrow, until final analysis or early stop.

Most data challenges occur during study execution, when patients are enrolled and followed up. At this stage, data is collected from the investigators and entered into a database, either converting paper-based forms or directly into an EDC system. Most trials involve multiple data sources (e.g., Case Report Forms, core lab tests, medical device data) which need to be integrated, cleaned and transformed into analysis datasets. For a more in-depth discussion, see *Data Challenges in Adaptive Trials*.[5]

4 R&D Senior Leadership Brief, The Adoption and Impact of Adaptive Trial Designs, Tufts Center for the Study of Drug Development, 2013.

5 C. Garutti, Data Challenges in Adaptive Trials. Programming Pharmaceutical Users Software Exchange conference (PhUSE), 2014.

The benefits of DMW's end-to-end data management capabilities are amplified in adaptive trials, where the data-cleaning effort is repeated multiple times for the same trial. Once set up, validation checks and transformations can be easily automated, so that each time new data is loaded in the clinical data repository, downstream operations are triggered and new derived datasets are updated in real time. The robust and flexible blinding capabilities in DMW ensure that the study team can perform its activities without breaking the blind, even when dealing with high number of interim looks.

The benefits of DMW's end-to-end data management capabilities are amplified in adaptive trials, where the data-cleaning effort is repeated multiple times for the same trial

Conclusion

Trial sponsors and CROs in today's environment face growing pressure to execute trials faster, with fewer resources and reduced cost – while still maintaining high-quality data for analysis, submission, and regulatory review.

This challenge is complicated by the increasing variety, velocity, and volume of data – sometimes referred to as "Big Data" – from multiple sources; the time and manual effort required to collect raw data, and clean and transform it for submission and analysis; and meeting regulatory requirements such as secured access, traceability, and blinding.

Increasingly, these life sciences companies are recognizing the limitations of EDC systems to solve these challenges, and are turning to Oracle Health Sciences Data Management Workbench to implement a true, integrated, end-to-end clinical data management platform across their organizations.

The result? Sponsors and CROs are:

- Saving time and cost by reducing time-consuming manual processes required to load, transform, and clean trial data from multiple sources
- Accelerating study setup and conduct by delivering cleaner data to biostats faster
- Increasing data quality with automated discrepancy reconciliation and data cleaning
- Providing better regulatory compliance with comprehensive security, audit trail, and two-way traceability across the discrepancy lifecycle

For additional information on Oracle Health Sciences Data Management Workbench, visit www.oracle.com/goto/dmw.

About Oracle Health Sciences

Backed by the resources of a Global 500 company, Oracle Health Sciences offers the industry's most comprehensive end-to-end clinical R&D platform supporting early to late-phase development, safety, and healthcare to help you reduce cost, risk, and time-to-market, while enhancing patient safety. With thousands of professionals in offices throughout North America, EMEA, and Asia, Oracle Health Sciences offers unmatched resources to enable your organization's goals today and in the future.

Oracle Corporation, World Headquarters
500 Oracle Parkway
Redwood Shores, CA 94065, USA

Worldwide Inquiries
Phone: +1.650.506.7000
Fax: +1.650.506.7200

CONNECT WITH US

oracle.com/healthsciences
healthsciences_ww_grp@oracle.com
youtube.com/user/oraclehealthsciences
facebook.com/oraclehealthsciences
twitter.com/oraclehealthsci

Integrated Cloud Applications & Platform Services