A PRACTICAL BUSINESS APPROACH TO DATA MANAGEMENT FOR CLINICAL TRIAL EFFICIENCY

Understanding the Problem

As clinical trials become more complex and expensive, pharmaceutical companies are continuously looking for ways to improve the efficiency of clinical trial operations. In addition to cost savings, well-planned and well-executed clinical trials can also deliver a substantial first-to-market advantage, especially in areas of high unmet medical need. Trial sponsors and contract research organizations (CROs) are looking for effective solutions to monitor, report and analyze the progress of clinical trials across many dimensions in order to make data-driven decisions and reduce the time and cost of drug development.

Even though all clinical data are now captured electronically, there are many disconnected systems being used to acquire the data and enable operational oversight and quality and safety monitoring: Interactive Response Technology (IRT), Electronic Data Capture (EDC) system, Clinical Trial Management System (CTMS), electronic Trial Master File (eTMF), quality management system (QMS), Safety, etc. In a typical clinical trial setting, each of these systems is likely to be supplied by a different vendor, built with a different technology and data model, managed by different business units and sometimes hosted externally to the sponsor. The project teams responsible for trial execution often have to navigate multiple systems and manually assemble the data in order to track study milestones, monitor site performance and generate reports for management and regulators.

Not only is this scenario inefficient, non-scalable and costly, it is also ineffective. It introduces delays and variability in information gathering and hence the decision-making process and can lead to human errors and gaps in trial oversight. So far, efforts to address operational reporting needs have fallen short, often due to vendor bias towards their systems or technology stack. Attempts to feed all relevant operational data into the CTMS have been limited by the CTMS design and the complexity of enhancing that design. Attempts to deploy federated reporting tools like Spotfire face challenges with data mapping and lack of consistent data access application programming interfaces (APIs). Adopting tools from a single vendor has not gained adoption due to substantial shortcomings in functionality, leaving the integration of best-of-breed products as the only viable option. There are several data warehousing solutions on the market designed to address this need, but none has gained widespread adoption to date.

The Challenge of Meeting Diverse Reporting Needs

It is also important to recognize that operational reporting needs are quite diverse across the clinical development cycle. The majority of needs can be satisfied with standardized reports on study performance vs plan (site activation, recruitment, milestones, etc.), unit performance (country, region, function, etc.), issue accrual and resolution (good clinical practice (GCP), serious adverse events (SAEs), protocols deviations, etc.), data quality (queries, source data verification, etc.), portfolio performance (financials) and CRO key performance indicators (KPIs). However, there will always be non-standard reporting requirements, such as trend analysis, root cause analysis, modeling and forecasting and others.

A successful solution for operational reporting must integrate well with market leaders in clinical data capture, and must also provide data integration options for niche data providers. It must address standardized reporting needs with a low-cost but effective solution and must provide options for ad hoc reporting. Additionally, it must be scalable and secure to meet the needs of a modern clinical development environment where multiple sponsors, CROs and data vendors often need to collaborate to ensure seamless trial execution and success.
Designing a Cohesive Approach

Our approach at Covance to clinical development analytics is based on the principle of separation of concerns, that is, the uncoupling of operational and clinical objectives. Operational objectives are focused on achieving optimal execution of clinical studies from a data quality, patient safety, timeline and cost perspective. In contrast, clinical objectives are focused on enrolling qualified patients, ensuring that the collected data are “fit for purpose,” and monitoring drug-related safety issues.

Following this core principle, we have developed two distinct data repositories as part of our Xcellerate® Informatics Suite: an operational data warehouse and a clinical data warehouse. The operational data warehouse is designed to support operational objectives and is based on a universal data model. Operational data are consistent across clinical studies and can be represented by a general data model that tracks site performance, subject visits and data quality. On the other hand, clinical objectives cannot be generalized and require an unconstrained data model. Covance has achieved both objectives by employing a traditional relational design for operational data and embracing an innovative, non-relational (so-called NoSQL) approach for clinical data.

There are several market offerings that claim to solve operational reporting needs through the use of Hadoop, NoSQL, data lakes and other “big data” approaches. We believe that the technologies themselves are sound, but their use for operational reporting is premature and misguided. The advantage of non-relational approaches such as data lakes is that they can ingest large amounts of data very efficiently while postponing data mapping/normalization until query time. While this is true in principle, the effort of data mapping/normalization still has to occur. Given that the majority of operational reports is standardized, it is more efficient to normalize the data upfront and thus simplify the effort at reporting time. Moreover, query and reporting capabilities for these technologies are still evolving and are not as powerful as the use of SQL in relational data stores.

Our solution for operational data integration and reporting consists of an operational data model, an operational data warehouse that implements that model, a set of extraction, transformation and loading (ETL) processes for loading operational data from data capture systems, a suite of web applications for standardized study and portfolio reporting and a set of APIs for ad hoc reporting.

Creating an Operational Data Model

Operational reporting requires integration and aggregation of operational data from multiple sources. The bulk of this information comes from a relatively small number of systems that are fairly standard in the types of data they capture. Virtually every modern clinical trial uses a CTMS to manage sites, site monitoring visits, site issues and trip reports; an Interactive Response Technology (IRT) system tracks subject enrollment, subject visits and drug allocation in a fairly standard way; an EDC system facilitates patient data collection, source data verification, and data query management; an eTMF tracks documentation compliance, and several others. Despite the variety of different solutions on the market, most of them are designed to capture very similar data and differentiate themselves in system capabilities, user interface and ease of configuration and management.

As a leading CRO, Covance has experience working with a large variety of data capture systems. In order to enable consistent and efficient study oversight, we have developed a unifying operational data model that represents operational entities, their attributes and their relationships. This model can support all of the standard reporting needs as well as ad hoc reporting requirements.
Our operational data model is developed to support both routine reporting and monitoring needs, as well as trend analytics. It is designed as a snowflake schema with shared dimensions representing key object attributes such as study, site, subject, contact, address, etc., and two kinds of fact tables: aggregated metrics for routine reporting, and individual events (e.g., payment activities) for ad hoc reporting and trend analysis. Wherever appropriate, we aggregate information at various levels such as subject, site, country, study and account, and we provide source-specific versions of metrics to improve data quality and support operational activities (e.g., IRT enrollment vs EDC enrollment).

Implementing a Data Integration Layer

Any comprehensive operational reporting solution needs to merge and aggregate data stemming from a multitude of activities and collected in disjointed systems, often hosted by different vendors. The main challenge with integrating operational data is the plethora of data capture systems currently in use, the proprietary customizations introduced in the various implementations of these systems and the scarcity or poor adoption of data standards. A successful data integration approach has to overcome challenges in data extraction (availability of APIs or direct access to the back-end database), data transfer (security and reliability), data validation (consistency and referential integrity) and data transformation (conformation, normalization and mapping).

Covance has adopted a data integration approach that successfully solves these challenges. We utilize the TIBCO Enterprise Service Bus (ESB) to orchestrate data extraction and transfer by invoking system API calls, scripting UI-driven operations, managing secure File Transfer Protocol (FTP) drop folders and handling messages. The extracted data are maintained as intermediate files for ease of handling and quality control, and are processed by source system specific data loaders. For example, extracts from an EDC system in Operational Data Model (ODM) format are handled by a specialized loader that understands the ODM data structure, performs specific validations, logs exceptions and filters out operational content such as page and query status. The IRT loader handles several typical formats of subject status reports and is configurable to accommodate differences between the commercial IRT systems in column naming and report layout.

CTMS, financial and master data sources typically allow direct access to their database schemas and can be handled with a traditional ETL approach and tooling, such as Informatica PowerCenter or SAP BusinessObjects. The same platform is used to perform most of the cross-source data mapping and to move data between landing, staging and data mart areas. While ETL tools provide powerful means for data transformation and validation, they have to execute transformation logic outside of the database. In some cases, especially when large volumes of data need to be handled, it is more efficient to embed the logic directly into the database in the form of a stored procedure.
Combining multiple modalities not only provides flexibility and allows our integration layer to combine data from a variety of sources in the most efficient way, it also enables us to achieve the optimum balance between ETL performance, speed of development, extensibility and maintainability. As new data sources emerge, they can be added quickly using the most suitable integration technique.

Configuring, scheduling and monitoring ETL execution is an essential component of data integration. All ETL processes are orchestrated and controlled via service database tables, and there are specialized user interfaces to configure ETL parameters and settings. ETL activity is also logged into the database tables, including both successful processing and exceptions. In addition to being logged, exceptions are communicated to the support team via email. Input data are subjected to validation checks for data contract compliance and referential integrity. Any data sets deemed non-compliant and any records that cannot be referenced successfully are rejected.

**Organizing the Operational Data Warehouse**

Our operational data model is implemented as a set of relational schemas in Microsoft SQL Server. The implementation includes a landing area where source data are brought in, a staging area where data are reconciled, harmonized and aggregated; and an Operational Data Warehouse (ODW) area where data are organized for consumption. The landing and staging areas are designed to support ETL processes and are never exposed to the end user. All data manipulations across the three areas are automated through validated ETL scripts, data loading utilities, stored procedures and configuration user interfaces (UIs).

The ODW is organized into subject areas around key data categories and data sources, and is designed as a snowflake, dimensional schema. Standard dimensions such as account, study, site, subject, trip, etc. are shared across the fact tables that aggregate the data in its latest snapshot as well as historically over time to provide a longitudinal view. In addition to the dimension and fact tables, a set of data views is available to support specific reporting, risk-based monitoring and outbound data transfer requirements. These data views provide additional data roll ups to support specific application needs, but also facilitate a “data contract” that shields applications from changes in the underlying data model. In fact, through versioning of data views, we can ensure continuing support of the existing applications while allowing our data model to evolve to meet new and emerging requirements.
Gathering Master Data

The ODW gathers a variety of operational data from a number of sources, including CTMS, EDC, IRT and other systems. Even before the data are brought into the landing area for processing, they are checked for referential integrity (data for unknown studies are rejected, records for unknown sites are flagged, etc.). This is necessary so that data from multiple sources can be successfully linked and the integrity and security of the data can be enforced. Therefore, master data on studies, sites, accounts, etc. is an essential data feed for ODW. The best source for master data is an enterprise master data management (MDM) system which drives standardization of key entities throughout the company. If a formal MDM feed is not available, a CTMS source can be used as a surrogate source.

Capturing Metrics & KPIs

Monitoring clinical trial performance involves periodic review of operational metrics and KPIs. Operational metrics are typically aggregated quantities derived from routine operational data and serve to identify risks, monitor and improve internal processes, manage resource allocation, strengthen investigator and sponsor relationships and other purposes. There is substantial variability in metric requirements depending on individual study needs, process differences, sponsor preferences and source system capabilities. Thus, a comprehensive solution for operational reporting must be able to support a variety of metrics and KPIs and allow flexible and expedient definition of additional metrics, as needed.

Figure 3: Representative study and portfolio metrics and attributes.
The majority of the operational metrics and KPIs available in the ODW are common across studies and programs, and are driven by the need for standardized study reporting and portfolio governance. Typical operational metrics reflecting site activation, subject enrollment, issue management, data collection, etc. are an integral part of our operational data model. Common roll ups and average values per subject, site, study, country etc. are pre-calculated as part of the standard ETL process, and are available for standard and ad hoc reporting through direct database access and APIs.

To support continuously evolving operational reporting and risk-based monitoring needs, Covance has developed a powerful and generalizable platform for calculating additional arbitrary metrics and KPIs as part of the daily ETL schedule. Aggregate metrics and derived quantities are defined using arbitrarily complex SQL expressions, can be derived from any data that is brought into ODW, organized as reusable and parametrized templates, and configured for study needs via an intuitive user interface. Furthermore, our platform allows us to easily configure the source-to-target mapping of the data fields that ultimately turn into metrics, making this the most robust, generalizable and extensible metrics architecture in the industry.

Enabling Data Access & APIs

Following best practices for service-oriented software design, we have developed a set of representational state transfer (REST) web services APIs to facilitate most common data access needs. RESTful APIs ensure secure data transmissions via HTTPS and bring additional benefits to guarantee performance, scalability, simplicity, modifiability, visibility, portability and reliability of reporting applications. We offer two sets of APIs: one designed to support standard reporting needs and one to support incremental and cumulative data extracts. Both sets of APIs can be utilized for ad hoc reporting and could be readily consumed by many commercial reporting tools, such as Power BI, Cognos, Spotfire, Tableau, Qlik, MicroStrategy, etc.

Figure 4: Metrics can be defined using arbitrarily complex and parameterized SQL statements.
Direct access to the ODW schema is allowed and facilitated via standard data access technologies and drivers supported by Microsoft SQL Server. Direct access to the database schema provides unlimited flexibility for ad hoc reporting and analytics through unrestricted access to the underlying data tables and the expressive power of the SQL query language. However, this approach does not allow for granular data access permissions and requires understanding of the data schema and sophisticated technical/programming skills. We reserve this approach for special analytics and data mining tasks.

Creating a Security Model

Our security model for data access is based on user roles and is enforced at the API level. Each user is assigned application specific roles that are enforced by application APIs and, hence, by the reporting application consuming these APIs. The granularity of user access is based on study and access roles such as clinical research associate (CRA), data manager, project manager, etc. User access policies can be used to automate study access and role assignment.

While the security model provides the access framework, managing user accounts can be taxing, especially if access needs to be disabled or removed quickly. All our reporting applications rely on federated authentication where user authentication is delegated to an identity provider supporting SAML 2.0. This approach enables single sign-on and delegates to the client the management of user accounts and the enforcement of password policies.

Leveraging Standard Reporting

The Covance Xcellerate suite includes web-based study and portfolio reporting dashboards that provide interactive, near real time views of operational data for tracking performance of ongoing clinical trials, monitoring emerging risks and enabling effective oversight, and facilitating project and portfolio governance. The dashboards rely solely on our operational data model, operational data warehouse and the corresponding APIs to supply all the data and metrics required. All the relevant operational data are brought into the data warehouse via the integration layer which includes connectors for data capture systems from leading vendors such as Oracle, Medidata, Almac, ERT, etc., but the design is extensible to accommodate additional data feeds.

Both study and portfolio dashboards have been developed to meet common operational reporting needs across many different types of studies. Our approach is based on the Pareto principle, also known as the 80/20 rule: we believe that about 80% of the operational reporting needs can be satisfied with a standard set of metrics and visuals. The remaining 20% fall into the category of ad hoc reporting, custom analytics and data mining, which would always require some customization or development effort. However, as the common reporting needs evolve, we are committed to incorporate additional data sources, metrics and visualization into our standard reporting application.

Study Dashboard

The Xcellerate study reporting dashboard is designed to enable project teams to track the progress of their trials against milestones and performance targets. The interactive dashboard offers longitudinal views of key performance indicators and metrics at the individual study level, with extensive drill-down, filtering and sorting capabilities. The dashboard is organized into several views, each reporting key metrics related to the performance of clinical development functions, sites, geographical regions or combinations thereof. The following views are provided and described in detail below: summary, site startup and enrollment, site performance, country performance, protocol deviations, data management and grant payments.

The study summary view provides visibility into whether key performance objectives such as site activation, subject enrollment and study completion dates are on track. It also includes key milestones and projections for meeting study objectives given the observed trends.
The site startup and subject enrollment view provides up-to-date information along with monthly trends on the number of sites activated, the number of site activation, initiation and close-out visits, the number of subjects screened, enrolled, terminated and completed, and many others. The view provides filtering capabilities by country and region, and allows for cumulative or incremental displays. The dashboard also provides so-called S-curve charts that show the cumulative number of sites that have reached a particular milestone compared against the initial and current plan (e.g., institutional review board approval, contract execution, green light approval, site initiation visit, etc.).

The site performance view provides information on site status, site subject enrollment and key site milestones. The view is fully interactive and allows filtering and sorting of sites by any of the metrics shown. Sites can be filtered and compared by geographic location, investigator, monitor, activation date or any other metric available. Several temporal trend views are also provided and help the user focus on recent developments.

For assessing country performance, we compare country milestones, site activation, subject enrollment and screen failure rates versus target. A world map allows geographic mapping of any desirable metric to enable quick detection of regional patterns. The view can also be sorted by any metric to identify leaders and laggards.

The protocol deviation view provides insights into protocol deviation accrual for study management and regulatory reporting purposes. All protocol deviations are classified by importance, category and resolution status. Each protocol deviation can be traced to the site and subject with extensive drill-down to any level of resolution, and details around protocol deviations can be exported as a CSV file for further processing.

Additional views offer insights into data management metrics, such as number of pages submitted, percent of source document verification, number of outstanding queries, query ageing, data entry timelines, etc. As with other views, the information can be dissected by country, query time and site. Bubble charts provide visualizations of data collection and data quality trends by country and region.
Portfolio Dashboard

Recognizing that portfolio governance and project management have very different reporting needs, we are currently developing a highly configurable, web-based portfolio dashboard designed to provide a comprehensive view of the health of a portfolio of studies at any desirable level of aggregation (therapeutic area, clinical indication, phase or any arbitrary grouping of studies selected by the user, as long as they have the appropriate access rights.). The portfolio dashboard uses a generalized, extensible framework for KPI, milestone and cycle time reporting, allowing users to view the status on key deliverables, past performance on completed activities, current position and risk of activities not yet completed, proposed key deliverables, geographic footprint overlaid with site performance and other startup, data management, financial and quality metrics. It is based on a generic attribute model and features a user interface that supports configurable interactive reports composed of multiple figures and tables, consistent attribute filtering and drill-down across all views, and ad hoc analytics based on a customizable charting library.

Supporting Ad Hoc Reporting

The Xcellerate Informatics Suite comes with a comprehensive set of RESTful web services APIs that allow programmatic access to all the data and metadata stored in the operational data warehouse. Additional reporting needs that are beyond the scope of the standard reporting UIs can be implemented using standard business intelligence (BI) tools like Spotfire, Tableau, Qlik, Power BI, Business Objects, etc. Most modern BI tools are able to consume RESTful APIs, retrieve the data of interest, perform additional aggregation and filtering and deliver it to the end users in a desirable visual representation.

Exploratory data mining and ad hoc analytics of the operational data is supported through direct access to the ODW schema. Popular analytics tools such as R and SAS can query the underlying SQL Server database directly using custom SQL queries. More sophisticated ad hoc reporting/analytic needs can be further facilitated through the development of custom data marts.

Designing for Scalability

Almost every person in a clinical development organization needs access to operational reports; hence, a successful reporting solution needs to be engineered for scalability. The Xcellerate architecture can support virtually any number of users by applying the principles of replication, load balancing and efficient engineering. Our operational data warehouse is based on Microsoft SQL Server 2014 and is populated by robust ETL processes. Depending on access needs, we can replicate multiple instances of the database behind a load balancing hardware such as F5. Multiple instances can be geographically distributed across our data centers in the U.S., Europe and Asia to eliminate bandwidth and latency bottlenecks, if needed.
As mentioned above, our reporting user interfaces utilize RESTful APIs to provide data with client-side rendering. This architectural pattern can be easily scaled to provide optimal performance to any number of users by employing load balancing and providing additional virtual servers to host the API layer. The design of the APIs has been carefully considered to minimize bandwidth and uses asynchronous calls to ensure that global latency does not adversely impact performance for remote users. Caching is also implemented at multiple layers of the architecture, from the database to the API to the web client, in order to minimize network bandwidth consumption and load on the data stores.

Realizing the Benefits

Clinical trials are very expensive, and unnecessary data quality issues or delays in trial execution can lead to significant loss of revenue for the sponsor companies. To accommodate a broad range of operational oversight and analytics needs, both from a sponsor and CRO perspective, we created the Xcellerate solution for operational data integration and reporting. Our goal in developing Xcellerate was to ensure that the design, planning and conduct of clinical trials proceeds in the most informed, efficient and effective manner possible. We believe that convenient, timely, integrated and contextualized access to all study data is an essential requirement for achieving this goal, and that scalable data integration, intelligent data engineering and tailored user interfaces are essential for driving user adoption and realizing the full value of our solution.

The suite of Xcellerate user interfaces are web-based, can be accessed easily and securely from anywhere in the world, and were designed and pressure-tested by subject matter experts with extensive experience in clinical development and clinical operations. Our operational data warehouse and data model can support a wide range of reporting and analytics needs and are extensible by design while new data sources can be incorporated quickly and with minimum effort with our flexible data integration architecture. Working with Covance, you can access an informatics suite that offers all the necessary features of a robust solution, including:

- Reliable and scalable architecture based on a proven technology stack and approach
- Comprehensive operational data model
- Scalable data integration framework with many pre-built connectors
- Extensible and configurable design simplifying introduction of new data sources and metrics
- Intuitive, powerful and elegant user interfaces and reporting capabilities
- Unrestricted support for ad hoc reporting and data access
- Secure, scalable and performant solution for enterprise-scale use

Following our mission to help clients bring the miracles of medicine to market sooner, Xcellerate is deployed internally to support the entire portfolio of studies managed by Covance. In fact, all Covance clients get access to the standard operational reporting capabilities as part of our full-service offering. Recognizing our sponsors’ desire to have comprehensive and consistent access to all clinical trial data across all their studies regardless of the delivery model (in-house or outsourced) and/or clinical partners employed, we offer Xcellerate as a standalone software-as-a-service (SaaS), independent of our traditional clinical development and central laboratory services. The Xcellerate Informatics Suite addresses an important unmet need in an area that has been historically stagnant and underserved by serving as a significant technological advance in the design, execution and monitoring of clinical trials.

REDUCE THE COMPLEXITY OF CLINICAL TRIAL DATA MANAGEMENT

Learn more about our drug development solutions at www.covance.com/Xcellerate