

# DATA GOVERNANCE IN THE CLINICAL TRIAL ECOSYSTEM

***MANAGING DATA ASSETS FOR MORE EFFICIENT DRUG DEVELOPMENT***

*Jaime Cook, Vice President, Technical Delivery*



## **Abstract**

*Data are the lifeblood of the drug development process. Expanding volumes of data, multiple data formats and dependence on an increasing number of eClinical systems make data governance essential for the efficient management of data assets across the research and development value chain. Good data governance delivers significant competitive advantage. A high-functioning clinical ecosystem drives better decision-making, operational efficiencies to reduce time and cost, and regulatory compliance to avoid costly errors and rework. This paper discusses the principles of data governance and how they are used to build a business intelligence framework that advances data quality, acquisition, and integration to deliver actionable information for use across the drug development enterprise.*

## **Managing the Complexity in the Research Ecosystem**

For biopharmaceutical sponsors, clinical trial data are both the greatest organizational asset and the greatest challenge. While clinical data drive commercial success, data volume, diversity, capture and analysis pose huge challenges. Efficient study execution and good decision-making depend on the transformation of clinical data into research intelligence—actionable information that informs study operations and culminates in successful regulatory submissions.

Automated eClinical tools have advanced this process, increasing speed and accuracy in trial management. But eClinical systems can also be part of the problem. Most are inflexible and incompatible with each other. They create disparate silos of data. Silo-ed data streams from multiple sources make it difficult to manage data across research processes. As the number and variety of eClinical tools increases, so does the risk of inconsistency, error and inefficiency.

Good data governance can help sponsors solve current problems in this complex technical environment and build a high-functioning data ecosystem to quickly adopt new data sources and methodologies, including the rapidly advancing mobile health (mHealth) technologies now enabling remote data collection.



## Principles of Data Governance

Data governance aligns people, processes and information technology to optimize the use and value of data across a business enterprise. This formal practice helps sponsors collect, integrate and analyze data strategically to advance their drug development programs.

Data governance underpins a framework in which new types and larger volumes of data can be harnessed to improve trial design and gain deeper scientific insights. It structures the data environment to facilitate real-time visibility into study operations—common views and analyses that enable effective collaboration, faster decision-making, and streamlined clinical operations.

**Definition.** Data governance is the overall management of the availability, usability, integrity and security of data used in an enterprise. Effective data governance maps an overall strategy and builds a framework that directs data management, distribution, protection, and alignment with industry specific regulations. Data governance defines and directs:

- Strategies for data collection
- Data standards
- Methods to support data integration
- Management of enterprise information.

**Goals.** The output of a successful data governance program is a high-functioning clinical trial ecosystem in which data are standardized and organized to: 1) promote more efficient and timely data access across stakeholders, and 2) enhance usability of information to achieve deeper insight into research processes. The ultimate goal is to achieve competitive advantage by harnessing data to drive time and cost efficiencies and increase the likelihood of successful trials.

**Processes.** To plan and implement a well-managed clinical trial ecosystem, data governance uses a centralized, top-down process to create a data environment in which all research stakeholders operate under a single framework that spans the entire drug development



process. A data governance oversight board plans and implements technologies and methodologies that:

- Standardize data management processes
- Create a centralized hub to promote collaboration
- Adopt open standards to maintain flexibility and scalability
- Provide tools for fast access to data assets and visibility into research processes

The company-wide governance framework is championed at the executive level to ensure compliance across operations and eClinical tools, including electronic data capture (EDC), interactive response technology (IRT), and electronic clinical outcomes assessment (eCOA), among others. Design and implementation of the framework is a long-term initiative, requiring commitment at all levels of the organization, among cross-functional stakeholders. The framework should promote joint ownership and accountability across departments.

The enterprise framework is built on the four pillars of data governance: data quality, acquisition, integration and consumption. These pillars are discussed in the following sections, using real-world illustrative examples.

## **Data Quality: Connecting through Standards**

Consistent data standards are necessary to underpin data quality, management, and applications across increasingly complex research processes. Failure to establish standards upfront makes it difficult—and in some cases, impossible—to connect data and systems for efficient study execution.

A common pitfall, for example, is the disconnect between information entered in laboratory notebooks and their use in a trial. These free text entry fields often have no relationship to fields established for entry inputs into other downstream systems. Valuable information becomes inaccessible or requires rework to connect it to related systems, wasting research time and money.



Effective standards also drive access to data *across* trials, providing insight into trial design and operations based on past research experience. With appropriate standards in place, data can be linked moving backward in time, much the way a genealogy traces ancestor lives. Standards make it possible to connect and trace previous research intelligence to mine historic trial data from the “genealogy” of a drug development program or therapeutic indication.

The work of the Clinical Data Interchange Standards Consortium (CDISC) has made notable progress in creating platform-independent, shareable and end-to-end data standards for clinical and nonclinical research. To date, seven foundational standards focus on core principles of data standard definitions and include models, domains and specifications for data representation. Standards focus on how to structure the data; not how data should be collected. Clinical Data Acquisition Standards Harmonization (CDASH) establishes a standard way to collect data in a similar way across studies and sponsors so that data collection formats and structures provide clear traceability of submission data into the Study Data Tabulation Model (SDTM), and in turn, more transparency for regulators. Continued global adoption of harmonized data standards requires collaboration across regulatory agencies, research sponsors, CROs, technology vendors and academia. (*source: <https://www.cdisc.org/standards/foundational>*)

***A Case of Lost Genealogy.*** A recent data quality assessment conducted by a major pharmaceutical company illustrates problems that can arise from a lack of pre-established standards. The sponsor was faced not only with data quality issues but also lost access to a valuable research genealogy.

In the sponsor’s pharmaceutical science laboratories, data were managed by a combination of paper notebooks, a laboratory information management system (LIMS), a chromatography data system (CDS), a scientific data management system (SDMS), and a materials assessment system (MAPP). The labs had recently adopted an electronic laboratory notebook system (ELN), which became the key system for creating, using, tracking, and storing experimental data. The labs also used vendor-provided systems for excipient data, drug product data and project codes.



The assessment analyzed metadata from the LIMS, CDS and ELN systems and their linkage to key supply chain systems and found numerous data quality issues: lack of consistent standards across systems, lack of quality measures, inconsistent data entry procedures and lack of system integration. A key recommendation was to establish a broad data governance program to advance data quality and usefulness.

The sponsor had intended to apply preclinical data from past work in another therapeutic area to streamline a program to develop 15 compounds. Assessment of the LIMS, CDS and ELN systems confirmed that no linkage was possible to give the sponsor access to this previous work. Without consistent standards and no uniform view across systems, most data could not be leveraged for learning.

Beyond identifying and defining standards, a multi-disciplinary process improvement initiative was required before the sponsor could begin its original goal of linking existing data for rediscovery efforts. This involved migrating and mapping, training people to implement standards, processes to ensure compliance with the standards and implementing governance structures to ensure value capture.

### **Data Acquisition: Managing More**

As big data reshapes drug development processes, sponsors must be able to manage more data, from more disparate sources, across more electronic information systems.

Novel sources include finance and business data, which can be leveraged from their silo-ed systems to support research. The emergence of mHealth technologies impacts both the type and volume of data as remote data collection takes clinical trials out of investigational clinics and into real-world settings. Sponsors will gain access to new types of real-world assessment, especially patient-focused eCOA. mHealth capabilities for continuous data collection and reporting will generate unprecedented volumes of data to be structured and analyzed.

Linking multiple systems and implementing new technologies pose increasing demands on existing research ecosystems. Data governance defines sources and types of data and designs strategies to access them. It establishes a framework to support data access from



multiple sources and systems, to relate data across systems, and to manage huge volumes of data without loss of quality or efficiency.

***A Case of Overload.*** This sponsor, a major global pharmaceutical company, was managing a large number of eClinical and operational systems. A new web-based application was implemented to serve as the principal clinical trial management system (CTMS) for study planning and tracking conducted by different business units. This global system provided web-based data entry for trial data.

As data volume increased, the sponsor was not able to scale up efficiently. Interfaces across the web-based CTMS and other eClinical tools in the enterprise system broke down under the demands of more data using antiquated and inflexible technologies. The effort to maintain these interfaces was very expensive, and the sponsor commissioned an assessment to address the problem.

The data acquisition assessment analyzed inbound and some outbound interfaces for the web-based system in order to design a strategy that would improve interfaces and reduce costs. A long-term strategy was developed to address the company's future integration needs using a flexible architecture that would allow the sponsor to scale and adapt to changes cheaply and easily.

## **Data Integrations: Connecting Silos of eClinical Data**

When all data assets are stored in one place, users have access to a "single source of truth"—a comprehensive warehouse of information that can be viewed, shared and analyzed to track study operations and respond to problems quickly. Data governance guides the process of integrating multiple, diverse data streams to create a central repository for all clinical and operational data. Additional types of data—like financial and business information—may be integrated as well.

Integrating clinical data is often a major bottleneck in clinical trials, especially in study startup where delays in patient enrollment and fulfillment of regulatory requirements are major contributors to cost overruns. Using traditional approaches, integration requires complex IT architecture and countless hours of mapping, cross-platform testing, and data



transfer validation. Data integrations typically cost hundreds of thousands of dollars and several months of development time for a given trial.

Newer cloud-based infrastructure is evolving as a viable means to centralize large volumes of clinical data. They are flexible and scalable, and they can include real-time open architecture to connect silos of clinical and operational data.

Powered by its comprehensive, connected data, the centralized repository becomes the hub of clinical trial operations with the addition of analytics and reporting tools.

***A Case of Data Traffic Jams.*** A sponsor needed to improve integration between an IRT system and a vendor's proprietary distribution system with the company's planning, manufacturing and distribution system.

For any given project, the sponsor worked with 1-2 CROs, multiple vendors, and hundreds of sites. The infrastructure required to manage these external systems was outdated.

Twenty-four integration endpoints, all of which triggered by events that took place in the IRT or vendor's distribution system, were connected by point-to-point interfaces, which posed a big risk to data integrity, speed and productivity. If one transfer failed, data flows for every connected system were affected. Even errors within an acceptable range caused a data traffic jam or worse, a snowball effect. For future studies, the sponsor wanted to support bulk drug distribution, which involved multi-layered file formats, and the capability to handle blinded kit types.

The solution involved an integration platform that would directly integrate and standardize data flow processes between systems, eliminating the need for data transfers and custom programming. Soon after completion, the integration platform was expanded to support for multiple studies.

The platform now enables faster data corrections through active monitoring and self-service error remediation. The new platform ensures that errors don't sit in a log. Instead, they are tracked to observe resolution. Data-driven actions can now resolve future problems instantly. Use of a cloud-based architecture offers the flexibility to add modules to the core engine and scale up as data volumes increase.



## Data Consumption: Analytics, Dashboards, Reports

Data consumption is concerned with optimizing the ways data are used. In the drug development enterprise, sophisticated analytics and reporting tools can turn a centralized data repository into a dynamic research platform that drives clinical trial insights and efficiencies.

These advanced integrated platforms give researchers real-time views and analyses of ongoing trial operations on digital dashboards. Role-based reporting offers detailed data views for key stakeholders, from study and program managers, to medical reviewers and senior management. Data are combined from multiple systems to provide a single accurate picture of trial events in real time; dashboards can show progress and events by site and even by one patient. Analyses and dashboards can be adapted for a given trial.

The result is visible, actionable study intelligence that can be used to track startup operations, conduct risk-based clinical monitoring, and enable adaptive trial designs. Data are combined, analyzed and displayed to track and improve operations including:

- Site selection
- Patient enrollment
- Site activation
- Clinical monitoring and risk mitigation
- Safety monitoring

***A Case of Overwork.*** Automated data platforms that combine, analyze and report trial data in real time eliminate errors that arise from manual processes and dramatically reduce workload and time. As data volume increases, lack of integration and automation makes reporting a daunting task.

Reporting became virtually unmanageable for a sponsor relying on manual processes to generate weekly comprehensive patient profile reports. Two high-level clinical operations staff would run reports from each of the company's multiple systems—eCOA, IRT, CTMS, and



laboratory systems—and load these data streams into Microsoft Excel. On average, it took eight hours to manually generate massive spreadsheets to combine, compare and report all the data. Over all, it took more than 20 hours a month to create a report that often was outdated before it could be completed.

Once the impact of dated reporting and wasted resources was evident to clinical operations management, the organization invested in data aggregation and reporting technology to present a patient profile dashboard in real time throughout the course of a study.

## **The Data-driven, Automated Future of Clinical Research**

Central repositories featuring analytic tools and dashboards are fast becoming the operating platforms of clinical trials. Such platforms are already offered by CROs and specialty providers to support conduct of sponsors' studies.

Data governance defines the quality standards, acquisition, integrations, and consumption of data that make these comprehensive, automated platforms possible. They provide competitive advantage by improving:

- Organizational efficiencies through better decision-making
- Operational efficiencies through cross-functional speed and insights
- Risk management to avoid time and cost overruns
- Regulatory compliance to streamline submissions and approvals

A range of tools exist today which allow organizations of all sizes to implement a cost-effective data integration platform in a cloud environment to connect the many sources of eClinical data. This eliminates the need to build and maintain costly integration infrastructures, broadening access to small and virtual companies.



## Building an Enterprise Framework

**Oversight Organization.** The first step toward implementing a data governance framework is to establish an oversight board of key information technology leaders and data stakeholders. Oversight board leadership includes five principal roles, shown in Figure 1.

**Executive Sponsor:** Serves as enterprise process owner; champions and oversees the data governance program at the executive level. (An organization's Chief Information Officer, Head of Information Management or Head of Data Management or similar position may serve as an executive sponsor).

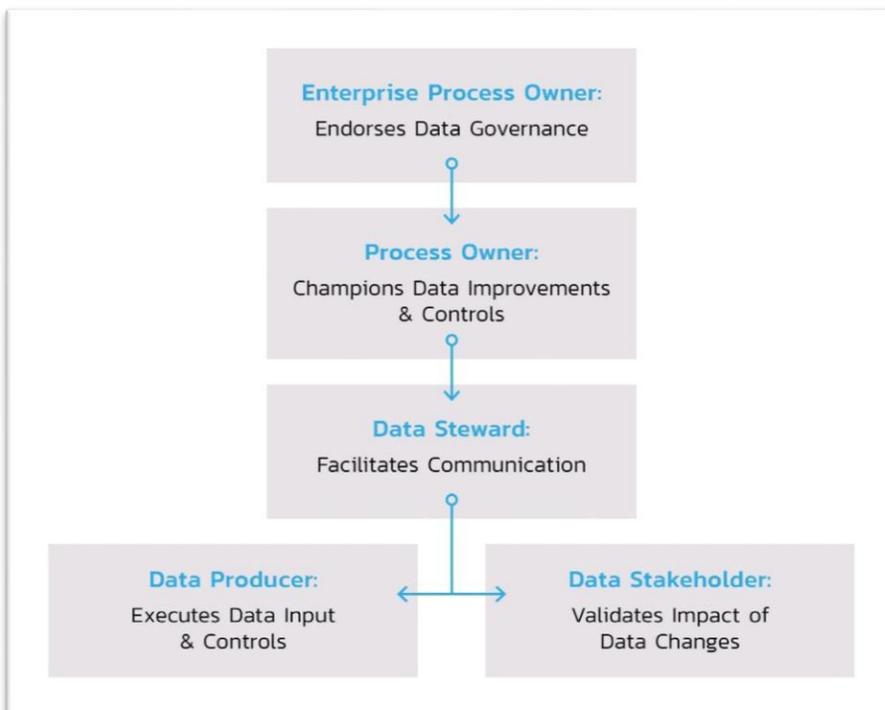
**Process Owner:** Directs the process to build the data governance framework; collects metrics, reports results, supports a universal data approach and educates the extended team on appropriate data entry. Process owners typically have data ownership roles and may be part of the organization's data management team or serve as a CTMS head.

**Data Stewards:** Representative group of data stakeholders across the clinical trial ecosystem; set policy, standards, data quality rules. Data stewards are typically comprised of data experts and day-to-day end-users.

**Data Producers:** Create, protect, control and distribute data to the Data Stakeholders. Data Producers can be anyone who access data on a day-to-day basis.

**Data Stakeholders:** Participants in conduct of the clinical trial, including the sponsor, clinical service provider, investigators and sites, patients, laboratories, technology providers, and other third-party vendors. Stakeholders scrutinize, apply and act upon data outputs and changes.

Figure 1. Data Governance Oversight Organization



**Implementation Roadmap.** One of the first tasks of the oversight board is to map the data governance workflow in three implementation tracks: user requirements; data and technologies; and solution architecture. A typical roadmap is shown in Figure 2.

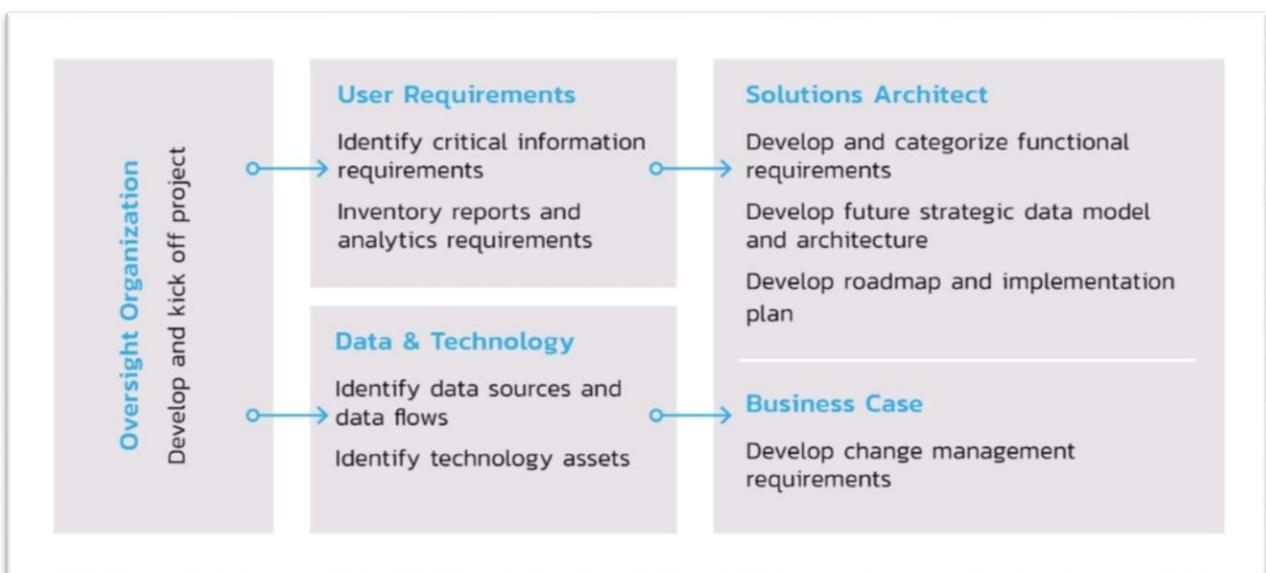
**User requirements.** Work includes defining mission-critical data requirements, inventorying key reports, and determining analytic requirements.

**Data and technologies.** This track focuses on identifying current data sources and high-level data flows. The business intelligence and technology environment inventoried, and current and planned data initiatives are documented.

**Solution architecture.** With the input delineating user requirements, data sources and technologies, the work to implement the framework architecture begins. Data requirements are organized and prioritized into subject areas and modeled strategically, first to create a

business intelligence strategy and then to develop “future state” architecture. This architecture guides the design of the data governance organizational structure and management.

Figure 2. Data Governance Implementation Roadmap





## **Conclusion**

In spite of strong consensus on the need for new approaches to management of its data assets, the biopharmaceutical industry remains slow to act. It must learn from disruptive innovation taking place in other industries to create value and much-needed efficiencies across research and development processes. Sponsors need novel, highly efficient approaches to quickly absorb, analyze and act on insights extracted from large volumes of data.

Industry adoption of initiatives, best practices, and technologies to create efficiencies, eliminate redundancies and reduce cycle timelines across the clinical trial ecosystem are slowly taking shape. While implementation challenges for both large and small organizations remain, many existing large-scale initiatives offer transforming effects on the way clinical development is conducted:

### **CDISC**

Industry-wide adoption of CDISC standards will expedite integration of electronic medical records with clinical trials to greatly enhance the speed, efficiency and safety of novel therapeutic treatments. New insights can be generated more quickly through mining of EMR data, observational studies may be conducted more rapidly, and clinical trial recruitment and conduct could be dramatically improved. Adoption of specific standards such as pharmacogenomics can contribute to overcoming barriers that impede advances in precision medicine, or personalizing prescribed therapies based upon a patient's specific set of biomarkers.

### **Cloud-based reporting**

Cloud technologies represent the next phase of data standards. As standards are defined for how data are stored and represented in the cloud, and HIPPA concerns are addressed, more industry providers and sponsors will adopt cloud-based reporting, replacing the in-house systems that many providers rely on today.



### **Data-driven decisions**

Powerful analytics that perform machine learning functions are transforming the clinical trial process through its ability to detect and explore outliers, trends and outcomes. Volumes of operational data can be analyzed across a range of scenarios, to reduce redundancy or add predictive insights into site-level performance, patient-level responses, and trial outcomes.

### **Data sharing**

The greatest scientific breakthroughs occur when the research community collaborates. The formatting of data to enable sharing can significantly shorten development timelines. Redundancy of effort will be significantly reduced when scientists and researchers can share what has worked and what has not.

Ultimately, organizations who invest in optimized R & D infrastructures, adopt business practices that involve standardized processes, and embrace new technologies that eradicate data silos and facilitate collaboration with other stakeholders will be best positioned for agility and efficiency, and for responding to future information needs, as they emerge. Today's competitive advantage will be tomorrow's essential operating components.