From Data Chaos to Order & Insight: Creating a Connected Data Ecosystem to Drive Drug Development

A PRECISION BRIEF

PRECISION quartzbio℠

From Data Chaos to Order & Insight: Creating a Connected Data Ecosystem to Drive Drug Development

Advancements in technologies that support precision medicine development efforts have led to an exponential increase in the volume, variety, and velocity of data generated across preclinical and clinical settings, including data available in the public domain. Unfortunately, these data sets are collected and generated from a myriad of sources and typically remain disconnected due to their complexity and disparate nature.

This white paper explores the challenges of organizing these complex data sets and makes recommendations for creating a data ecosystem that can generate actionable insights.

Today, disjointed data leads to a resource-intensive approach to analysis, consisting of ad hoc efforts to compile the data needed to test a particular hypothesis or answer a specific scientific question, such as, “Is my therapy engaging the intended target?” But science is unpredictable and delivering on the promise of precision medicine requires the ability to interrogate and analyze data across all available data sets. Our vision of the future is one where data analysis is proactive and continuous: Data is collected, harmonized, and integrated in real time and made accessible for visual, computational, or AI-driven interrogation.

Making the transition from disjointed data to data asset requires sponsors to systematically connect information silos to create a unified data ecosystem. This enables researchers, investigators, and clinicians to dive deeper into the “why” and develop increasingly effective and innovative treatments based on richer, more comprehensive insight generation. Successfully completing this shift can create massive scientific value.

Big Data in Precision Medicine—Sample, Data, Insight

Drug development is driven by the generation and collection of proprietary data (measurements) and analyzing this information to create knowledge or insights (results or conclusions from experiments).

Sponsor data is at the heart of drug development and includes preclinical, specialty lab, clinical sample metadata, and clinical subject data. Biomarker data plays an especially important role across the discovery and development continuum, including target identification, disease understanding, characterization of mechanism of action, predictive and surrogate biomarkers, and patient stratification.

However, the proliferation of biomarker data poses a significant challenge when compiling and integrating data sets comprising tens of millions of data points from multiple labs covering diverse assay types.

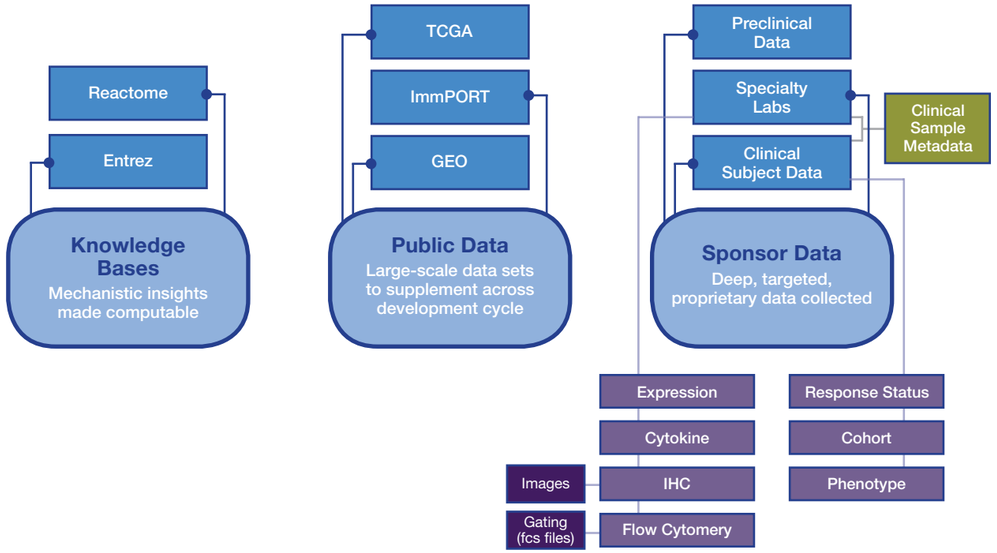
In addition to sponsor data, the use of public data sets is increasing. For example, the Cancer Genome Atlas (TCGA) is a federally funded effort that has captured multiomic measurements from >11,000 patients representing more than 30 cancer types. Data repositories such as TCGA afford researchers access to data to enable more focused preclinical and clinical work that would otherwise be financially unviable or take years to generate.

The proliferation of biomarker data poses a significant challenge when compiling and integrating data sets comprising tens of millions of data points from multiple labs covering diverse assay types.

Knowledge comprises mechanistic insights such as published and inferred biological pathways or molecular interactions derived from data and literature. Often, this knowledge is unstructured and may reside in documents, Powerpoint presentations, or even the heads of researchers. Public knowledge bases attempt to extract results from published papers and aggregate them into a format more accessible for researchers. For example, Entrez Gene collates information about official gene names, aliases, and annotations about gene function. In addition, numerous sources, including KEGG, Wikipathways, or Biocarta, provide information about the interactions and functions of biological signaling pathways.

Harmonizing and organizing these complex data sets within the context of prior knowledge—in other words, creating the unified data ecosystem—is critical for generating actionable insights.

Figure 1. The Current State of Data Chaos



Breaking Down the Data Silos

There are significant challenges to systematically integrating data collected at different times and across different therapeutic areas. For instance, at a tactical level, biomarker data in clinical trials is typically generated by multiple specialty labs

At a strategic level, data silos may lead to missed opportunities.

that may separately run immune system profiling, gene expression, and cytokine assays. All of this data must be linked to gain visibility into the trial. These data linkages further require integration of clinical metadata that maps back to samples, patients, and cohorts to inform decisions such as dose escalation.

At a strategic level, data silos may lead to missed opportunities. Consider a scenario where prior knowledge data sets are used for clinical development planning, and then new biomarker data is generated in a clinical study in one therapeutic area.

Currently, these 2 data sets are rarely combined to inform future development decisions in that therapeutic area, and they are virtually never used for informed decision-making across indications. In fact, establishing connections between these sets may be valuable for mechanism of action (MOA) analyses or patient selection.

In order to fully realize the benefits of preclinical, clinical, and public data sets, a shift in thinking is required. Rather than simply leveraging data to address targeted needs during drug development, in this new era of the data ecosystem each data point contributes to a larger data platform, where analysis can be completed in a context that is broader than a single clinical trial or indication.

Value of a Data Ecosystem

At an enterprise level, the creation of a data ecosystem has a multitude of benefits across therapeutic areas including the ability to:

- Capture information that may be critical to understanding why a drug does or does not work in certain patients
  - In cases where an end point is reached, sponsors experience a significant value inflection if they have a comprehensive data package that includes publicly available data
  - In scenarios where an end point is not met, sponsors may still have clinically relevant data that enables them to adapt their programs
- Learn new information about the disease and drug under investigation—insights valuable for informing clinical development planning
- Validate the strength of evidence indicated by preclinical models
- Identify potential opportunities for combination treatments
- Repurpose existing compounds for new indications
- Pivot programs based on real-time data and timely insight generation
- Build strategic data and knowledge assets that not only add value to ongoing clinical trials, but also link back to drug discovery

Ultimately, the objective of creating a data ecosystem is to make every bit of data count, thereby decreasing discovery and development risk. This involves moving from a disjointed data approach, where data analysis is conducted in a reactive ad hoc manner, to a data asset mindset where each piece of data is an asset that permits insight generation and value creation. Given the broad implications a data ecosystem has on an organization within and across trials—and the effort involved in establishing one—the transition to a data ecosystem requires executive-level commitment and a C-suite champion.

A Data Asset

Creating a data ecosystem involves moving from a disjointed, ad hoc approach to data to a mindset where each piece of data is viewed as an asset that permits insight generation and value creation

Making Data Ecosystems a Reality

To bring the vision of a data ecosystem to life, sponsors need to first de-silo data and make it usable for a variety of scientists with diverse technical backgrounds. The next step is defining and executing a comprehensive approach to quality checking, integrating and processing the data in its entirety.

A comprehensive approach to data integration must include the following:

- **Method for storing all data in a retrievable format**
- **Process for transforming data into usable formats**
- **Procedure for ensuring data quality**
- **Audit trail**
- **Capability for reporting that allows for immediate action if needed**

Once data is assembled, knowledge and insights must be captured and linked to the data ecosystem. This knowledge contextualizes the data and provides opportunities for early hypothesis validation or rule out, as well as feature reduction. Knowledge should be stored and managed in a clean, harmonized, and machine-readable format that enables rapid and consistent retrieval. It also must be easy to curate to further minimize ad hoc and ongoing researcher burden.

The final piece of the data ecosystem puzzle is a flexible, learning-based computational framework. Given the complexity of the problems sponsors are trying to solve, there is no one-size-fits-all solution or silver bullet. Advanced platform components that integrate data-, mechanistic-, and knowledge-driven approaches or hybrid approaches are needed to make interrogation of the data ecosystem accessible to a larger number of scientists while reducing reliance on pure correlation.

Of course, building the necessary components to support a data ecosystem requires focused commitment from a dedicated team. One critical challenge is assembling the necessary cross-functional mix of big data and biological expertise—from experimental design, data collection, and data management to data analysis and eventual translation back to the bench.

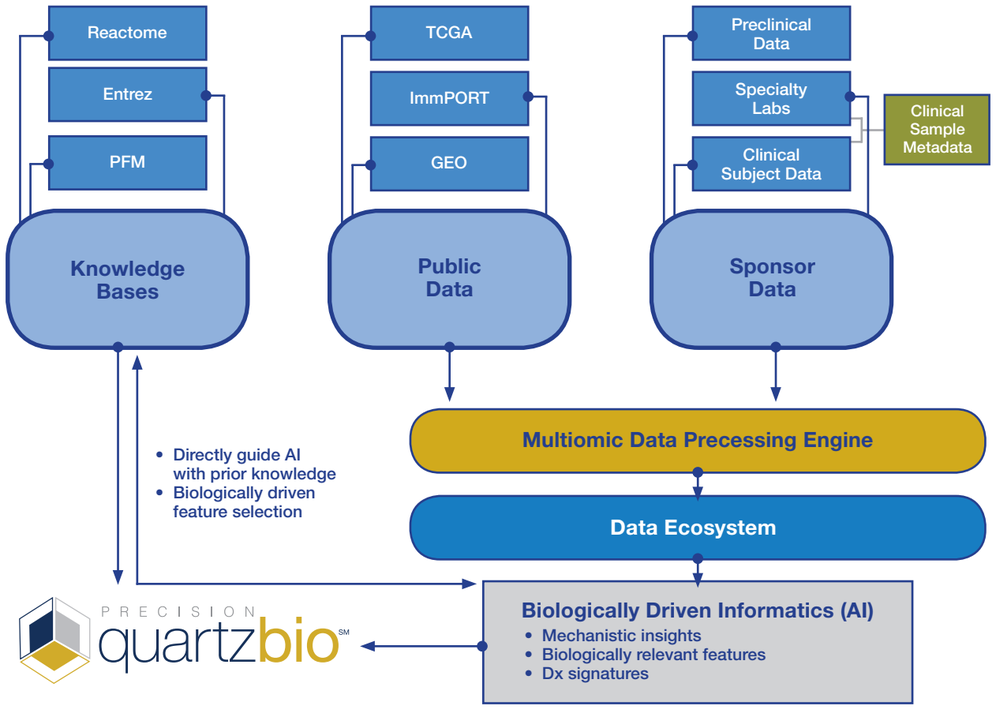
QuartzBio: A Cloud-Based Multiomic Data Integration and Informatics Platform

QuartzBio takes a systematic approach to making clinical data ecosystems a reality by combining an industry-leading multiomic data integration and informatics platform with deep multidisciplinary subject matter expertise. QuartzBio integrates diverse skill sets spanning software and operations engineers, data architects, knowledge management experts, translational biologists, clinicians, and data scientists who are trained in the peculiarities of molecular data. The platform is engineered specifically for biomarker and multiomic data management, data delivery, visualization, and on-the-fly analytics and reporting, and includes a suite of tools allowing sponsors to query and export data.

With QuartzBio, organizations can access their complete multiomic (eg, genomic, transcriptomic, proteomic, flow cytometry) data in real time to inform on-trial decisions and enhance overall organizational knowledge. This data is combined with publicly available data and knowledge bases to advance the understanding of disease mechanisms, drug target discovery, advanced biomarker identification, patient stratification, and drug positioning—ultimately resulting in accelerated drug development, more effective treatments, and improved patient outcomes.

Figure 2. QuartzBio Data Ecosystem

With QuartzBio, sponsors can integrate public data and knowledge in one place, enabling access and deep interrogation of the clinical data ecosystem.



Conclusion

As data grows in quantity and complexity, sponsors need a proactive approach to both manage and extract value from it. Ultimately, the goal is to tease out the clues hidden within the wealth of biomedical and clinical information that exists—and continues to be generated—to find paths that advance diagnostic and clinical outcomes for patients.

To increase the pace of innovation in precision medicine, well-curated, high-quality data needs to be ubiquitous and universally accessible within and across organizations. Making data available to the widest possible audience optimizes imagination, out-of-the-box thinking, and discovery. We expect that, just as electronic data capture has become the norm for clinical trial data management, data ecosystems will become the standard for unlocking the full potential of precision medicine.

PRECISION quartzbio℠

PRECISION℠ FOR MEDICINE

Visit us at: Quartz.Bio