WHITE PAPER

Big Data Is Driving the *Future* of Molecular Diagnostic Testing

A trusted research environment can help diagnostic companies meet changing expectations and create new revenue streams.

DNAnexus®

Introduction

Thanks to the era of big data in healthcare, diagnostic companies have a remarkable new opportunity to innovate — both to develop better evidence-based products for patient care and to leverage their aggregated data to evolve new lines of business.

At the same time, diagnostic businesses face more pressure than ever, including:

- Discovering and validating better biomarkers
- Getting support for reimbursement in a crowded marketplace
- Navigating a complex intellectual property landscape
- Commercializing assays as regulatory guidelines shift

Diagnostic companies can bring new business opportunities to fruition and support new data-driven products to improve patient care with innovative use of their growing data assets. But to do this, they need to securely and efficiently receive, integrate, analyze, share, and distribute sensitive patient data from multiple sources in a way that meets the needs of their customers.

What Is a Trusted Research Environment?

A trusted research environment, also called a TRE, provides sophisticated data management and analysis capabilities with the security and regulatory compliance required for protected or sensitive classes of data. TREs are often cloud-based in order to support collaboration, enabling multiple locations to access data in compliance with use restrictions through specific permission controls.

TREs are becoming increasingly specialized, especially systems designed for molecular and clinical data. This includes:

- Data warehousing features designed to handle the scale of variant-level queries
- Purpose-built analysis tools and application programming interfaces (APIs) designed for the life sciences sector
- Robust compliance with the latest industry security and compliance regulations

While an organization could invest in developing its own data environment, it is often more expedient in the near term and more cost effective in the long term to implement a third-party option. Commercially available TREs typically have a wider variety of sophisticated data analysis tools as well as fully-managed security protocols and regulatory compliance than many in-house, do-it-yourself options.

For most diagnostic companies, a third-party TRE can leapfrog the months or years it would take to build their own infrastructure. A commercially available TRE makes it possible to avoid drawn-out delays and unanticipated costs in rolling out their data strategy.

With the right platform, diagnostic companies can quickly deploy valuable offerings for their customers and partners. For example, they can responsibly license their de-identified test data to pharmaceutical partners or provide more sophisticated test interpretation and decision support for clinicians. By making a TRE part of their infrastructure — particularly one purpose-built for clinical-omic data — companies can greatly reduce time, cost, and risk of handling their partners' and customers' data, bypassing significant hurdles to their omic real-world data strategy.

Evolving Expectations

Diagnostic companies have participated in industry partnerships in the real-world data (RWD) research market for years by providing patient data for biopharmaceutical research. However, patient data, particularly genomic data, has grown in demand, as has emphasis on protecting privacy by controlling the use of such data. Delivering patient data in a TRE ensures the ability to meet privacy standards, such as readily reporting on data access and restricting data movement based on regulations and norms that continue to evolve.

At the same time, healthcare providers also have changing expectations. Just a few years ago, diagnostic companies were only responsible for giving healthcare providers reliable and validated assays with the basic information to interpret the data they generated. Times have changed, and quickly. Advanced data offerings from many of today's industry leaders provide analytics to improve interpretation of molecular features for more actionable results. At the same time, health systems have begun to expect access to their aggregated test data for their own mining and retrospective query and analysis. Increasingly, customers want to add their own clinical data on top of the molecular test data from their diagnostic test vendor, as well as to see their patients' data in the context of de-identified aggregate data to benchmark results.

Delivering such valuable capabilities requires a TRE with sophisticated data management features for user-friendly queries, as well as analysis tools for cohort comparisons and deep dives into patient-level data.

Trusted Research Environments to Help Drive Growth

A TRE can serve as the central foundation to enable a number of important growth drivers for diagnostic companies, all while allowing these companies to focus on their core competencies and differentiators instead of building off-the-shelf infrastructure.

GROWTH DRIVER 1:

Sharing Real-World Data While Protecting Patient Privacy

Many diagnostic companies have chosen to diversify and evolve their business by providing aggregate, de-identified data to pharmaceutical and biotech companies. In the era of precision medicine, biopharma scientists are eager for access to biomarker-related data, especially when it's based on real-world patient results. Such data can help hone patient stratification guidelines and give candidate therapies a better chance of succeeding in clinical trials.

Making data available through a TRE allows diagnostics companies to share data responsibly with their industry partners, while ensuring that protected data cannot be downloaded or shared, in compliance with best practices for security and control. Protecting patients' data is key to future-proofing a business where the regulatory framework around what and how to share is likely to shift in the coming years.

Controls in a TRE make it possible to grant different levels of access to certain people or roles. Importantly, a TRE with controls that are easily configurable enables a single data admin to make role-based access more or less restrictive. The admin can do this in compliance with evolving policies without having to rebuild infrastructure. In addition, TREs have a full audit trail for all data access that ensures a company's data administrator can readily show, whenever necessary, that patient privacy has not been compromised. More specifically, a cloud-based TRE can add value by providing a central location for collaboration, permitting access to trusted collaborators in different global locations to co-analyze and glean more insights from the data they are licensing. It is key that the TRE also include commonly used analysis tools for industry partners, as well as features to easily integrate user-preferred advanced analytic tools. Lastly, the TRE framework greatly reduces the manual workload typical of standard, FTP-style file-sharing protocols.

GROWTH DRIVER 2:

Expanding Access to Outcomes Data

Diagnostic companies that have access to rich, real-world clinical and outcomes data for patients receiving their tests are at a competitive advantage. They are able to:

- More quickly amass real-world evidence (RWE) to support reimbursement
- Glean insights to improve their diagnostic products
- Support a more valuable commercial RWD offering
- Have deeper market knowledge about how their products are used

Responsibly incentivizing data sharing by partner health systems is one approach to obtaining clinical outcomes data. For this approach, it is common to either create a registry or request direct electronic medical record (EMR) access in exchange for linking the clinical outcomes data to patient test results and assay data. The linked data is then returned to the health system partner in a structured, standardized format. This can make the test data more valuable to the ordering institution, especially when this enhanced data is delivered with access to advanced analysis tools.

CASE STUDY: MYRIAD GENETICS

Myriad Genetics sought to provide a community resource to advance patient care by combining structured clinical data with genetic and genomic results. It would allow researchers to collaboratively analyze and understand patterns to help improve and advance cancer care. Myriad realized that building the infrastructure internally would be too costly in labor and time. The company sought a partner with a proven, ready-to-use platform that could provide speed-to-market advantages. With DNAnexus, Myriad launched the Myriad Collaborative Research Registry, also known as the MCRR portal, on the DNAnexus precision health data platform. DNAnexus was not only able to provide a ready-to-go, secure, compliant, and highly scalable environment for Myriad's clinico-genomic data, but also to act as a key data design partner with an institutional understanding of what providers and researchers need to do with the information.

Today, the Myriad Collaborative Research Registry includes data on more than 1 million patients, making it one of the largest pan-cancer registries available for research; it will support transparent clinical data sharing to advance the field. The registry will continue to expand over time and is expected to include patient outcomes, responses to treatment, and additional tools to further enhance research.

This mutually beneficial "give to get" model of data exchange can help improve patient care by supporting health systems in growing their data compendium with each test. The data can then be mined and analyzed for quality improvement or to inform treatment decisions based on patterns over time. In some cases, the data can be merged with de-identified aggregate data or with public data for richer insights across groups. At the same time, the diagnostic company is able to incorporate the outcomes data into their product development and innovation initiatives to ultimately better serve health systems' future patients.

For a diagnostics company to forge and support such partnerships, it needs a platform that not only is a secure place to house the health system's incoming data, but also provides a way to link and ingest the data and deliver it back to provider-side end-users in an analysis-ready environment. TRE capabilities become a crucial component to ensure the delivery of protected data back to multiple institutions without the risk that users from different institutions could inadvertently access one another's protected data.

CASE STUDY: CITY OF HOPE

City of Hope is one of the largest cancer research and treatment organizations in the United States. As a National Cancer Institute-designated comprehensive cancer center with sites across the nation, City of Hope has access to a trove of de-identified clinico-omic data from patients through its EMR and public data sources.

City of Hope partnered with DNAnexus to create the POSEIDON platform to empower investigators across the City of Hope network of more than 35 hospitals and treatment centers with access to multi-omic, precision medicine data. When selecting a platform partner, City of Hope required the ability to import data from its EMR, to quickly scale, and to be capable of supporting a multitude of data modalities. It also required a platform that is easy to use for both bioinformaticians and oncologists, with secure access controls to protect patient data. More than just a data repository, POSEIDON has helped inform City of Hope's unique in-house drug development program, support translational research projects, help place patients into clinical trials, and enable collaborative data-driven clinical decisions on complex cases.

Recognized internationally with the Bio-IT World Innovative Practices Award in 2023, POSEIDON's impact also extends to inclusion in significant grants, more than 100 peer-reviewed publications, international partnerships, and keynote presentations, solidifying its position as a transformative force in advancing healthcare research and innovation.

GROWTH DRIVER 3:

Building a Competitive Data Advantage

In addition to the two previous growth drivers, it is becoming increasingly common for molecular diagnostics companies to deliver a single patient's test results in the context of other patients' de-identified test data to improve interpretation. A company can distinguish its offering with a variety of approaches, including with its own specialized analytics and decision-support tools delivered on a secure platform. These may include developing methods to flag variants based on patterns across populations, or tools for providers to make "patient like mine" queries in clinical care pathways.

For a diagnostics company seeking to distinguish itself and add value to test results by delivering aggregate data (or its derivative), a TRE is key to retaining patients' and providers' trust. Even when de-identified, the aggregate data should only be shared with highly controlled access to protect both patients' and companies' interests. A TRE solution should not only link and make the data accessible for analysis, but also easily support integration of the company's interpretation tools.

Key Considerations

While many TREs are available, they don't all offer the same capabilities. When considering the best option, diagnostic companies should pay particular attention to the following factors.

Designed for life sciences: Consider a precision health data platform specifically designed with the life sciences sector in mind. It should be purpose-built to provide access for other researchers and clinicians in a private analysis workspace with built-in, industry-standard data science tools and omics workflows. It should be able to handle large-scale omics data volumes, as well as linked, structured file data. It should also handle clinical data extracted from EMRs for analysis.

Security: Given the sensitive nature of the data involved, security must be a top priority. All TREs offer some degree of security, but only the best TREs incorporate the measures required for working with academic medical centers and other healthcare institutions. Look for an option that features robust end-to-end encryption, full audit trails, penetration testing, remediation, robust monitoring, and other security protocols.

Regulatory compliance: In the healthcare field, security is only half of the data-protection story. The other half is covered by regulations. Any TRE worth evaluating must comply with the following clinical regulations:

- HIPAA
- CAP/CLIA
- 21 CFR Part 11
- GDPR
- IVDR
- FedRAMP
- ISO 270001
- PCI DSS
- Data Privacy Framework
- Cyber Essentials/Cyber Essentials Plus

CASE STUDY:
UNITED KINGDOM BIOBANK

In 2020, DNAnexus was selected to be the sole provider of a cloud-based analysis solution for researchers to access and analyze UK Biobank's vast database, including its groundbreaking whole-genome sequencing data from 500,000 participants. The secure, cloud-based UK Biobank Research Analysis Platform, also known as UKB-RAP, allows approved researchers from around the world to use the data to drive the discovery of new diagnostics, treatments, and cures. More than 30,000 approved researchers from at least 90 countries have registered to use UK Biobank, the world's most comprehensive source of biomedical data. These scientists are provided with the tools and computing power to analyze a vast wealth of de-identified health and lifestyle data.

TheUK Biobank Research Analysis Platform solution built by DNAnexus leverages the power and scalability of the DNAnexus platform to enable thousands of users around the world to securely access and analyze the phenotypic, genomic, metabolomic, imaging, proteomic and forthcoming single-cell RNA data from the 500,000-participant cohort. The protected cloud environment meets the most rigorous standards for data quality, security, privacy, and safety. The platform proactively manages local, regional, and international compliance requirements in order to protect participants' confidentiality while maximizing the opportunity to help researchers access and collaborate on this valuable data.

Configurability: Data administrators should be able to easily configure access to serve a broad range of users and quickly. This includes being able to:

- Configure access and controls at the user and organizational level
- Adapt to inevitable changes in privacy regulations and norms without having to rebuild infrastructure
- Configure access to restrict downloading, copying, sharing, and access to analytic tools

User-friendly interface: The optimal TRE will offer easy access to an intuitive and user-friendly interface with configurable controls. It's also worth considering that many users will require query and analytical tool options for both technical users, such as analyst coders, and non-technical subject matter experts.

Collaboration: TREs should support collaboration, enabling multiple locations to access data. The access should be in compliance with use restrictions through specific permission controls. It should also meet security and regulatory requirements for all regions where the data is accessible.

Analytical pipelines: Analysis capabilities really set the best TREs apart from all others, providing sophisticated functions that go well beyond housing and protecting data. The ideal TRE will have a suite of readily-available analytical tools and analytical workflows. It should also allow users to bring their own algorithms and analysis tools and to create custom workflows when needed for meaningful data interrogation.

Data management: Within a secure and compliant environment, users should have the ability to integrate and enrich existing data from multiple sources, and to present them so they are compliant with FAIR (findability, accessibility, interoperability, and reusability) guidelines. Omics data management also includes capturing metadata for optimal flexibility and utility.

Customization: When considering a third-party provider, find one that can deliver a white-labeled system that allows custom branding for a consistent user experience across diagnostic tests and the omics data management platform.

Take the first step to optimize your operations.

See how DNAnexus can help with your next project or data initiative.
info@dnanexus.com
dnanexus.com/clinical-diagnostics

ABOUT DNANEXUS

DNAnexus, the enterprise platform for precision health, is on a mission to accelerate the development, approval and delivery of personalized treatments. Building on 15 years of bioinformatics innovation and genomics expertise, DNAnexus provides the cloud platform that centralizes and enriches multimodal omics data, supports an extensive suite of informatics use cases, and allows secure collaboration across the care continuum. DNAnexus powers a connected ecosystem trusted by the world's precision health leaders. This flexible ecosystem makes omics and real-world data accessible, actionable, and secure, while unlocking insights that improve patient lives. For more information, visit www.dnanexus.com or follow @DNAnexus on social media.