

# MIT SMR CONNECTIONS

EXECUTIVE CONVERSATION



## Data-Driven Health Care: Enhancing Patient Outcomes Through Digital Engineering

COMMISSIONED BY:

 Persistent



**Nick Jena,**  
Vice President, Health Care and Life  
Sciences, Persistent Systems



**Madhuri Hegde,**  
Senior Vice President and  
Chief Scientific Officer,  
Revvity

Cause for celebration: Humankind has entered a new era for detecting disease.

Screening that identifies disease before any symptoms manifest enables early intervention and treatment. Millions of lives – of newborn babies and people of all ages – are being saved thanks to new screening techniques that employ a multidimensional approach, combining time-tested biochemical techniques (traditional testing measures such as blood tests and imaging) with omics technologies and other big-data sources that are now widely available. This breakthrough combination helps all of humanity, but individual people also benefit, from newborns whose autoimmune disorder is detected right after birth to adults who find out their lung cancer is susceptible to immunotherapy. These are exciting times for researchers, physicians, and patients, and for the population as a whole.

But that capability involves a precise balancing act: handling ever-increasing amounts of patient data, safeguarding its privacy, and ensuring that only authorized users can access it – while also remaining flexible enough to adapt to new discoveries and developments.

In this Executive Conversation, **Madhuri Hegde** of Revvity and **Nick Jena** of Persistent Systems discuss addressing those challenges in a process they describe as “a marriage of biology and digital engineering,” with the goal of making data more meaningful so that it translates into better patient outcomes. They also share insights about creating an effective data-engineering ecosystem, the roles of artificial intelligence (AI), machine learning (ML), automation, telemedicine, and other technologies to continuously improve ways to identify and treat disease.



## SETTING THE STAGE: LEVERAGING DATA FOR HEALTHIER HUMANITY

► **Madhuri Hegde:** Let's start with newborn screening. That's a proven way to save lives and improve outcomes. Revvity screens about 40 million infants annually for 29 disorders; we estimate that the process saves about 75 babies' lives every day. We work with both diagnostics in a clinical setting and with pharmaceutical providers to help in disease targeting.

With early intervention, which can be very simple, the individual might lead a healthy life with a normal life expectancy. For pediatric or neonatal conditions, the sooner these conditions are detected and treated, the better.

In fact, early detection starts from conception right through adulthood. Now you can do noninvasive prenatal testing to identify certain abnormalities. The ultrasound can also identify some structural abnormalities. But even before conception, a couple can get carrier testing to identify which recessive conditions they're carrying that could be inherited by the fetus.

We always pair genomic data with biochemical data and imaging data. But there are also "omics-based" data sets, where you're bringing in all the different modalities together under one umbrella to do clinical interpretation.

► **Nick Jena:** That kind of early detection requires being able to harness vast stores of patient data while making sure that it stays private and only authorized users can access it. That, in turn, requires having a rigorous data engineering framework. But the approach must still be flexible enough to accommodate changes in use over time as well as the inevitable increases in data volume.

The sheer amount and complexity of data generated by omics technologies require specialized computational resources for effective management and analysis. Integrating diverse omics and clinical data types and dealing with high dimensionality pose additional challenges. Data quality, standardization, and the need

for substantial computational infrastructure further complicate the process. To address all that, you need to be able to analyze, visualize, and manage omics on a scalable cloud platform.

## A MULTIDIMENSIONAL APPROACH TO UNDERSTANDING DISEASE

► **Hegde:** Two or three decades ago, early detection and intervention depended largely on non-automated tools, basic bench tools that researchers used to identify, quantify, and treat disease. But in the last two decades, that technology has changed significantly. We still use traditional techniques as confirmatory methods, but now we take a multidimensional approach based on new technologies. Automation has also changed how we do things in clinical laboratories.

Researchers, scientists, and clinicians are all getting excited, and rightfully so, about using AI and ML in screening and diagnosis. Here's a simple example of how AI will revolutionize how we operate: The human genome was sequenced in 2003, and at the time, we expected the cost of genome sequencing for an individual would be about \$1,000. Now that cost is just \$200. But we like to say *interpreting* that genome is the \$1 billion question.

**“Early detection requires being able to harness vast stores of patient data while making sure that it stays private and only authorized users can access it. That, in turn, requires having a rigorous data engineering framework. But the approach must still be flexible enough to accommodate changes in use over time as well as the inevitable increases in data volume.”**

— Nick Jena,  
Persistent Systems





What that means is, it's not just about sequencing and the output of the sequencing; it's also about how we are going to understand what these results actually mean. How do we interpret what the genomic data is really telling us?

► **Jena:** Data engineers are bringing a lot of tools and solutions to the table to help answer that question. First of all, the vast amount of data generated by gene sequencing requires specialized computational resources. IT and digital engineering play a crucial role in providing the necessary infrastructure and tools to manage and analyze complex multi-omics, clinical, and health care data sets. They enable efficient data storage, processing, and sharing through high-performance computing clusters, cloud-based platforms, and big data technologies.

These advancements facilitate the integration and interoperability of diverse data sets, allowing researchers, clinicians, and other relevant users to combine and analyze data from different sources seamlessly.

Additionally, leveraging ML and AI algorithms empowers researchers to extract meaningful insights from vast amounts of data to uncover hidden patterns and associations that might not be apparent through traditional approaches. That early detection potential enables timely interventions and personalized treatment plans – and, ultimately, improving patient outcomes.

► **Hegde:** Many countries have population genome-sequencing projects today. That data, all over the world, provides invaluable clues to understanding genomic variation between populations and unlocking the identification and treatment of disease.

The challenge is to collate all that data, understand it, access it, and apply it to patient care to help an individual who needs treatment. With billions of sequences available, what does it mean when a physician is looking at an individual's data and trying to decide how to treat him or her? That's where the power of AI comes into play. What we can do today is get all that data under a single platform and apply the technologies behind AI to interpret it and translate it into patient care.

AI's ability to recognize images, for example, is important because the human brain can assimilate only so much. But recognizing and identifying those images and putting them in front of the person who's going to interpret them is key. From the human genome perspective, we are all about 99% or more the same. It's essential to know which differences are actually meaningful and could cause disease. Interpreting that is both art and science. AI really powers through this massive amount of data and presents it to individuals to then interpret it.

► **Jena:** Just to reiterate: The most important rule of disease management is early detection – the earlier the better. Advancements in technology, and particularly in data engineering, AI, and ML, can empower researchers to extract meaningful insights from vast amounts of data to uncover hidden patterns and associations that might not be apparent through traditional approaches. This potential for early detection potentially enables timely interventions and personalized treatment plans, and all that adds up to improved patient outcomes.

ML and AI algorithms empower the researchers to extract those insights, which wouldn't even be visible with traditional approaches. We have to uncover a vast amount of data. And there are so many hidden patterns and associations that might not even be apparent through traditional ways. These algorithms analyze huge amounts of information, such as patient data, which includes omics medical imaging and electronic medical records [EMRs], identifying patterns and markers. This can indicate the presence of disease even before the symptoms manifest.

► **Hegde:** But you have to approach all of this with caution. Take the example of a single DNA change seen in a patient today. If your AI is pulling data only from the Caucasian population, then we could designate the DNA change incorrectly. But if there's a broader data set, it could turn out this variation is very common in, say, the Asian population and, therefore, is not a disease-causing change.



AI empowers us to bring these data sets together. But the final call has to be made by humans: How we are going to interpret the data and use it in patient care?

For example, we had a case where a physician approached us in the last trimester of her own pregnancy. A clinical interpretation indicated that the fetus might be affected by a life-threatening disorder, and the family needed to decide whether to terminate the pregnancy.

At that time, our own database essentially told us to look at this particular change with caution: It might not be a disease-causing change because it's common in a certain small population in India. That analysis was possible because we had the algorithm in place already. The family elected to continue the pregnancy, and a healthy child was born.

Now, data can come in many different formats, and much intermediate translation is needed to make the data accessible. We started with simple Excel data sheets. As the data size started getting bigger and bigger and bigger, it became obvious that we needed to develop tools to handle these large data sets.

► **Jena:** We worked with Revvity to develop proprietary software based on an approach called the *ordered data interpretation network*. This undertaking demanded significant effort, especially considering the varied formats in which data may be presented, as well as the enormous volume.

The software takes the data Revvity generates and adeptly organizes it in meaningful ways for effective, efficient, and accurate clinical interpretation. Ultimately, it performs annotation and visualization of sequenced data samples collected from individual patients and large data sets. That enables researchers, clinicians, physicians, and scientists to visualize huge amounts of data in one frame based on different filter criteria as well as being able to easily conduct analysis and profiling of all variants.

Why is that important? Because most commercially available tests look at an extremely limited number of letters in a gene sequence. Those represent a very small percentage of a person's

**“Digital engineering helps early detection by enabling the use of the vast amounts of data generated by these processes. Another trend currently driving advancements is the growth of telemedicine and remote monitoring in disease screening. Those innovations will help these providers and researchers reach out to populations in remote areas.”**

— *Madhuri Hegde,*  
*Revvity*

overall disease risk and, for that reason, can provide a false sense of reassurance or concern. The proprietary software analyzes each of the 80 genes in its entirety and provides an in-depth assessment.

The platform I've described here is a web-based tool accessible to authorized users worldwide via VPN. At first, only a few directors and scientists used the tool, but we all knew the numbers would grow. We had to build in various levels of access around different roles with encryption and role-based access.

## DRUG DEVELOPMENT TARGET IDENTIFICATION

► **Hegde:** Now we are looking at how we can bridge the divide between diagnosis and life sciences tools to better support therapeutic interventions. AI and ML play pivotal roles in orchestrating these efforts.

Target identification is a process in which therapeutic developers can precisely determine which individuals within a patient population will benefit from a specific pharmaceutical agent. A



recent publication has shown that providing genetic evidence to the U.S. Food & Drug Administration in the approval process tends to expedite approvals for new pharmaceuticals.

Revvity provides omics-based, protein-centric methodologies, known as proteogenomics, for detection for the entire ecosystem of drug discovery from preclinical to clinical research. The tools are used for target discovery, starting with genome-wide analysis for identifying different DNA changes that contribute to disease. Certain individuals could be resistant, while others might be more susceptible to the drug agent. Genomics data can tell us that. RNA sequencing is the next level that can tell us about the expression of our genes, and that can aid in pharmaceutical development.

Another major advancement: We can now train our systems to comprehend the current treatments available for a particular ailment. There are 7,000 rare diseases. It's unrealistic to expect a physician in a busy clinic to instantaneously determine the best course of action for each malady. Revvity's proprietary software

**“Most commercially available tests look at an extremely limited number of letters in a gene sequence. Those represent a very small percentage of a person’s overall disease risk and, for that reason, can provide a false sense of reassurance or concern. The proprietary software analyzes each of the 80 genes in its entirety and provides an in-depth assessment.”**

— Nick Jena,  
*Persistent Systems*

system helps physicians by presenting additional information gathered through AI — things like new treatments and applicable clinical trials. That powerful capability saves physicians' time. Hours and days matter when critical disorders are at stake.

In addition, in clinical settings where patients are sick, information about results must be communicated clearly and accurately. A genetic counselor talks to physicians to help them understand what's in the clinical report. Telling a physician to tell families that, “The test has an indeterminate,” is possibly the scariest thing. No one wants to do that.

The genetic counselor's job is to explain the results and the necessary next steps, such as the testing of additional family members. That, in turn, can help determine whether a variant is inherited or *de novo* [that is, not seen in either parent]. This guides the laboratory in the next steps toward variant classification.

With healthy newborns, the screening program operates as a comprehensive ecosystem extending beyond the test itself. That includes the parents and the hospital staff responsible for collecting screening samples within 48 hours of birth and the lab responsible for generating and delivering the results to the parents. If the screening identifies a disorder, immediate medical attention for the baby is crucial.

In terms of developing the tools and software and applying AI, we need to do more to prevent false-positive results, which cause a lot of stress to patients and their families. We also need to continuously improve test sensitivity and specificity. Bringing accuracy to this entire engine is an evolving process. When a new gene is introduced to the newborn screening panel, a lot of thought goes into that.

Digital engineering helps early detection by enabling the use of the vast amounts of data generated by these processes. Another trend currently driving advancements, as Nick can confirm, is the growth of telemedicine and remote monitoring in disease screening. Those innovations will help these providers and researchers reach out to populations in remote areas. But the right tools and infrastructure are required.



► **Jena:** We are indeed seeing an emergence of telemedicine and remote-monitoring technologies in disease screening. These innovations enable health care providers to remotely monitor patients' conditions, provide consultations, and adjust treatment plans without the need for in-person visits. This trend improves access to screening services, particularly for underserved populations and those in remote areas.

IT and digital engineering play a crucial role in providing the necessary infrastructure and tools to manage and analyze complex multi-omics, clinical, and health care data sets. They enable efficient data storage, processing, and sharing through high-performance computing clusters, cloud-based platforms, and big data technologies.

## MANAGING MASSIVE DATA STORES

► **Hegde:** A typical clinical laboratory deals with massive amounts of data on a daily, weekly, or monthly basis. These data sets continue to expand exponentially. ML algorithms must be designed to mine these data sets and create solutions broadly throughout the health care industry. Communicating the accuracy of that data through that entire channel and ecosystem is absolutely critical.

The problem for clinical laboratories is that they have only so much money for data storage. And how much data should be stored and retrieved? We often hear that data storage is easy and cheap, but data retrieval still isn't so cost-effective today.

► **Jena:** At Persistent, discussions revolve around vast amounts of data, measured in terabytes and zettabytes — one trillion gigabytes — of data. It's conceivable that in the future, we might not retain every bit of data within the system. Instead, we may opt to preserve the meticulously extracted database. The crucial objective is to adopt a data-driven approach, emphasizing the establishment of a resilient framework for the effective management, processing, and seamless integration of data.

**“A typical clinical laboratory deals with massive amounts of data on a daily, weekly, or monthly basis. These data sets continue to expand exponentially. Machine learning algorithms must be designed to mine these data sets and create solutions broadly throughout the health care industry. Communicating the accuracy of that data through that entire channel and ecosystem is absolutely critical.**

— *Madhuri Hegde,*  
*Revvity*

Of course, data security and privacy are critical because you're dealing with patient information, which is sensitive and highly regulated. You need technology to help with the data encryption part. Normally in these systems, data gets transferred from one person to another. That could be from a physician to a director to someone from a government agency. Robust data encryption and access control are paramount.

Due to the volume of data, it's important to use automation wherever possible, especially for the de-identification of data in its native format. The system extracts the text and images from the native format. It can automatically remove PHI [protected health information] even from the digital-imaging side of things. That's important because sometimes the digital imaging has information that can't be transferred or shouldn't be exposed.

And, again, the system's design must be flexible to evolve as user needs change over time. Inevitably, more and more users will be added. When you are building these applications today, it could be a few users; tomorrow, thousands of users might use those applications. The system might go from being an application to being a platform, as it did in our proprietary software project with Revvity.



► **Hegde:** The technology continues to evolve, and the data sets keep getting larger and larger. It continues to be a dynamic process. So it's truly a marriage of biology and digital engineering.

It's astonishing, at times, to think how fast the field of early medical interventions has developed. Just a few decades ago, the techniques were completely different. We can work much faster and more accurately today.

Revvity presented an abstract for ultra-rapid genome sequencing to the American College of Medical Genetics and Genomics annual meeting. We envision being able to get a diagnosis for a sick baby in the NICU [neonatal intensive care unit] in under five days. That was unimaginable in 2005 or 2006. It almost feels like future fiction, but it's a reality. That's what advances in technology combined with AI and ML make possible today. But for individual babies and their parents, it's not an abstract research project. Early intervention is about saving precious lives.

► **Jena:** That's exactly why we do what we do at Persistent, with the goal of improving health care through solutions that revolutionize disease diagnostics and management. It's about transforming the fields of health care and life sciences, ultimately leading to better health outcomes for people worldwide.





**Nick Jena** is vice president of health care and life sciences at Persistent Systems. His team of subject matter experts focuses on high-value solutions and custom development for health care and life sciences by leveraging advanced technologies, including artificial intelligence, cloud computing, automation, and analytics. He has been closely associated with the health care and life sciences industries for more than seven years. His role involves evaluating and promoting the latest solutions, partnerships, and techniques in product engineering, with emphasis on those two industries.

**Madhuri Hegde** is the senior vice president and chief scientific officer of Revvity and an adjunct professor of pediatrics at Emory University School of Medicine. She has more than 25 years of experience in clinical diagnostics and is board-certified in clinical molecular genetics. She serves on the boards of the American College of Medical Genetics and Genomics Foundation and the American Board of Medical Genetics and Genomics, among others. She received bachelor's and master's degrees from the University of Bombay, India, and a Ph.D. from the University of Auckland, New Zealand, completing postdoctoral studies at Baylor College of Medicine.

### ABOUT PERSISTENT

With more than 23,800 employees located in 21 countries, Persistent Systems (BSE & NSE: PERSISTENT) is a global services and solutions company delivering digital engineering and enterprise modernization. As a participant in the United Nations Global Compact, Persistent is committed to aligning strategies and operations with universal principles on human rights, labor, environment, and anti-corruption, as well as taking actions that advance societal goals. With 268% growth since 2020, Persistent is the fastest-growing Indian IT services brand, according to Brand Finance. For more information, please visit [www.persistent.com](http://www.persistent.com).

### ABOUT REVVITY

Revvity (NYSE: RVTY) provides health science solutions, technologies, expertise, and services that deliver complete workflows from discovery to development, and diagnosis to cure. Revvity is revolutionizing what's possible in health care, with specialized focus areas in translational multi-omics technologies, biomarker identification, imaging, prediction, screening, detection and diagnosis, informatics, and more. With 2023 revenues of more than \$2.7 billion and more than 11,000 employees, Revvity serves customers in pharmaceutical and biotech companies, diagnostic labs, academia, and governments in more than 190 countries. For more information, please visit [www.revvity.com](http://www.revvity.com).

COMMISSIONED BY:



**MIT SMR Connections** develops content in collaboration with our sponsors. We operate independently of the MIT Sloan Management Review editorial group. Copyright © Massachusetts Institute of Technology, 2024. All rights reserved.