

# [Interview] A New Way To Work With Data In Life Sciences

Sept. 25, 2020 by Andrii Buvailo

Founded by renowned database researcher, Turing Award laureate MIT Professor Michael Stonebraker, Paradigm4 is not just any data analytics company in the Life Sciences. The organization is built on the decades of pioneering research in database design and possesses unique technological know-how in scientific data management and scalable computation.

The firm has recently launched its REVEAL™: Single Cell app to offer biopharmaceutical developers the ability to break through the data wrangling and programming challenges associated with the analysis of large-scale, single-cell datasets.

Paradigm4's Agile Science Platform, REVEAL™, transforms hypothesis generation and validation to advance drug discovery, biomarkers, and precision medicine, providing FAIR (Findable, Accessible, Interoperable, and Reusable) data access and elastically scalable analytics and machine learning to power discovery from population-scale to n-of-1.

Below is my interview with **Marilyn Matz, CEO and Co-founder at Paradigm4**, and 2020 NACD Directorship 100™ honoree, and her colleague **Dr. Zachary Pitluk, VP of Life Sciences and Health Care at Paradigm4**.

**Andrii: How did Paradigm4 start? What was the key driver behind the founder's ambition to delve into a complex area of scientific data analytics?**



Marilyn:

Michael Stonebraker, a Turing Award Laureate who has been behind almost all major advances in databases for over 30 years, spent time listening to scientists talk about the limitations of current technology for scientific data management and scalable scientific computing. He created SciDB to meet their requirements for an array-native scientific computing platform that let them focus on their science without getting bogged down with computer science implementation details. We co-founded Paradigm4 to build out the technology developed in Mike's lab at Massachusetts Institute of Technology (MIT) into a robust software product. This is Mike's eighth company. Our key driver was to transform the pace of daily research and create a platform that provides a new way for researchers to integrate, share and gain insights from multidimensional data to help them to make scientific breakthroughs. We support our customers to achieve more efficient hypothesis testing and evidence generation to advance drug discovery and precision medicine.

**Zachary:** We have created one of the world's most scalable and adaptable growth platforms for scientific and biomedical data. Through that, we're enabling the seamless connection of multidimensional data from public and proprietary sources so that scientists can ask more questions, and get more meaningful answers, quicker than ever before.

---

**Andrii: Paradigm4 team builds a data analytics platform for the life sciences -- can you briefly describe the solution you offer? What key components does your platform include?**

**Marilyn:** Our Agile Science data engine, SciDB™, is purpose-built to handle large-scale multidimensional scientific data. It is a next-generation system, with storage organized around multidimensional arrays and vectors, that enables computation and sophisticated data modeling. It is a unified, enterprise-ready storage and elastic computing engine – a massively parallel, transaction-safe, array-oriented, analytics solution for high impact translational science.

SciDB rapidly serves up the selected data of interest because it preserves the logical structure and co-locality of data in its native data storage format. That is what makes SciDB a scientific data management engine at its core and differentiates it. It runs analytics, signal processing, image processing, and machine-learning elastically in parallel on data distributed across a cluster; on any cloud or on-premises.



**Zachary:**

Scientists told us they wanted higher-level, use-case focused solutions, not just workspaces which require



them to assemble their own management systems. We responded to that with REVEAL™, our suite of apps that power discovery from population-scale to n-of-1. Each is designed for scientists and bio-informaticians to ask relevant research questions without the need for programming or expert IT knowledge. REVEAL™ Agile Science apps serve up multidimensional data with ease, so that researchers can focus on the science.

The REVEAL™: Single Cell app offers biopharmaceutical developers the ability to break through the data wrangling and programming challenges associated with the analysis of large-scale, single-cell datasets. We also have the REVEAL™: Biobank app, which brings together multiple data types, such as multi-omics data; practitioner, hospital, diagnostic codes and prescription history; as well as biometric and imaging data to support scientists in cohort creation for population-scale translational medicine and healthcare research.

## **Andrii: What are the typical use cases for the biotech and healthcare clients?**

**Marilyn:** We are actively working with leading biopharma companies globally as well as research institutes. One of our current projects is working with Alnylam Pharmaceuticals to expedite their research on one of the biggest genetic projects ever undertaken – the UK Biobank. Over 500,000 people have donated their genotypes, phenotypes, and medical records. With so much data available on such a large scale, Alnylam's scientists had a challenge on their hands when it came to extracting meaningful information from it and making valuable connections that could unlock breakthroughs in scientific research. They chose to work with Paradigm4 because of our experience working with complex data on such a massive scale for storage, rapid access, and computing.

The UK Biobank captures genomics, longitudinal medical information, and images (from MRI scans etc), so having all that data in one place allows researchers to correlate someone's traits and presence/absence of a disease, or even susceptibility to diseases like COVID-19, with their genetic make-up. Alnylam has used our REVEAL™: Biobank app to help use these correlations to investigate causes of disease and identify potential treatments.

Our goal was to provide the users at Alnylam with an API where they can quickly connect to a system and start a task – for example, looking at a diagnosis of a condition like heart failure or high cholesterol in a sub population and selecting certain variants of high impact i.e. only selecting the genetic changes that

have the highest probability of causing the disease. Alnylam's scientists can then run computationally-intensive GWASs to investigate the connections between the genomes and the phenotypic data – and get those results quickly, e.g. running 1 billion linear regressions in less than an hour in very low cost using spot instances. They've published multiple papers and posters, results enabled by Paradigm4's agile science platform.

**Andrii: The market of artificial intelligence (AI)-driven data analytics tools and platforms in the Life Sciences is expanding fast, with already more than 230 biotech-focused companies developing or applying analytics or predictions models in drug discovery and clinical research. How is your product unique? What are the key attributes that persuade clients to choose your solution over other existing ones?**

**Marilyn:** With our Agile Science solution, data is 'science-ready' – we are providing full application solutions that provide higher level, more intuitive interfaces. Ultimately, we want to allow users to ask and answer questions faster and in a more interactive way with less coding effort.

**Zachary:** All AI methods rely on cleaned and organized data. By streamlining the whole process, making it more efficient and allowing higher throughput, we can help companies reduce the resources dedicated to data computation and therefore reduce costs. Our solutions are future-ready and equipped to cope with challenges related to data volume, diversity, scalability, and new algorithms.

**Marilyn:** Importantly, we are partners, not vendors, so we will help users to access and use their data intuitively and get answers, fast. Our team is made up of experts with backgrounds in bioinformatics, biochemistry, applied mathematics and scalable parallel computing. It is that broad knowledge base across both the scientific and computing industries that means we are best equipped to help solve our customers' data challenges.

**Andrii: Do you have any projects related to COVID-19 research? If so, can you explain how your product is contributing to a global fight against coronavirus?**

**Zachary:** Recently, in partnership with a leading pharmaceutical company [1], we used the REVEAL™: Single Cell app to analyse cells in the COVID Cell Atlas – a database of cells which stores information from patients infected with COVID-19. REVEAL™: Single Cell evaluated the expression of key SARS-CoV-2 entry associated genes and queried the current database (2.2. million cells, 32 projects) to obtain the results in <60 seconds. We highlighted that cells expressing COVID-19 associated genes are expressed on multiple tissue types which, in part, helps to explain the multi-organ involvement in infected patients observed worldwide during the pandemic.

The REVEAL™: Single Cell database was used as a reference to ask questions relevant to drug development and precision medicine regarding cell type and co-expression for genes that encode proteins necessary for SARS-CoV-2 to enter and reproduce in cells. We have been able to demonstrate that the app enables quick profiling of key genes involved in COVID-19 and supports additional use cases that require evaluation across a large database of single cell expression datasets such as vaccine candidates for infectious diseases, biomarkers for oncology patient stratification, and immunology-related disorders. The paper is available in preprint now. However we'll be sure to talk more about this study once the paper has been published along with our ongoing Single Cell work, so keep an eye on our website for more information.

## References:

Kumar, N, Golhar, R, Sharma, KS, Holloway, J, Sarangi, S, Neuhaus, I, Walsh, A, Pitluk, Z (2020) Rapid single cell evaluation of human disease and disorder targets using REVEAL: SingleCell™ <https://www.biorxiv.org/content/10.1101/2020.06.24.169730v1> (Unpublished – manuscript pending publication)

- Paradigm4