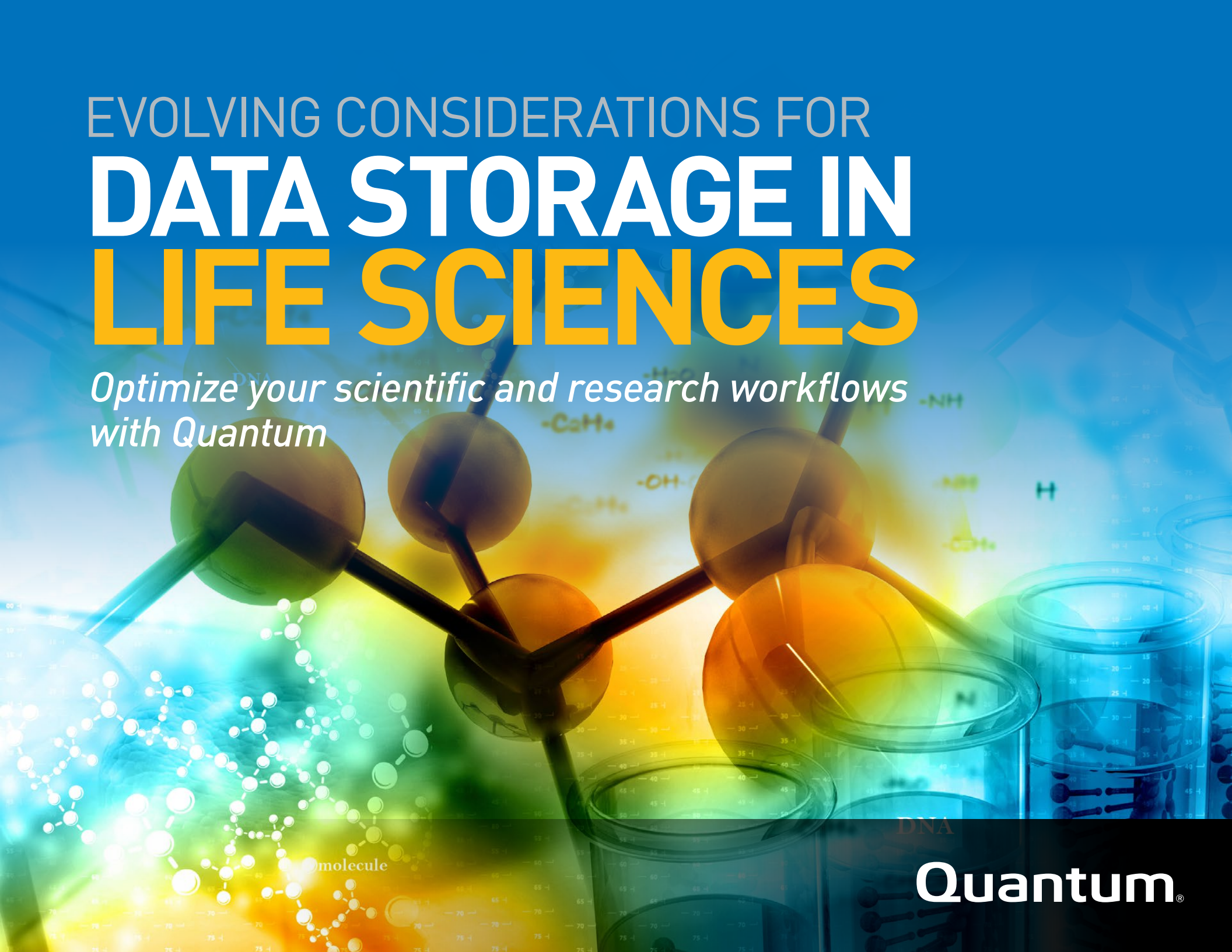
EVOLVING CONSIDERATIONS FOR

# DATA STORAGE IN LIFE SCIENCES

*Optimize your scientific and research workflows with Quantum*



Quantum®

SUPPORTING

# NEW, INNOVATIVE LAB TECHNOLOGIES

Quantum

# The latest lab instruments are driving the need for powerful IT resources in the life sciences.

Laboratory technologies are evolving rapidly. The latest instruments for life science research enable scientists to visualize highly detailed 3D models of protein molecules and assemble genomes in a fraction of the time they could in the past. These instruments are helping to accelerate the journey toward personalized medicine and deepen the exploration of new treatments for cancer and infectious diseases.

New lab instruments are also producing tremendous amounts of data. For example, in the field of genomic research, next-generation sequencing (NGS) instruments can output 1 TB of data per hour.

Cryo-electron microscopy (cryo-EM) is another burgeoning field that can produce massive data volumes. With cryo-EM, researchers study protein molecules at cryogenic temperatures, taking numerous images of each molecule and constructing 3D models. Researchers then use those models to study cancer at a molecular level and develop strategies for combating emerging viruses, such as Zika. The cryo-EM workflow can generate 5 TB of data in 24 hours.

Life science teams often require substantial high-performance computing (HPC) resources to process and analyze data created by new instruments and techniques. In fact, at some research institutions, life science departments consume more compute and storage resources than the traditional data-driven sciences of computational chemistry, astrophysics and climate research.

Now, many organizations need to bolster their storage environment. They need solutions that allow them to ingest a large amount of raw data from instruments, present data for analyses, preserve all data for the long term and ensure that data remains readily accessible by collaborative research teams. These IT solutions should not be overly complicated to manage: scientists should remain focused on their scientific research, not IT.

IDENTIFYING

# EMERGING IT CHALLENGES


Quantum

# Growing data volumes and disconnected systems threaten to slow research.

The development of new scientific instruments and emergence of new techniques are creating important research opportunities for life science organizations. To take full advantage of those opportunities, however, organizations must address several key IT challenges:

## DATA GROWTH

Today's scientific instruments are generating exponentially more data than instruments from just five years ago. They are part of workflows that incorporate a wide variety of file types, from text and binary files to databases and directories, in sizes ranging from a few kilobytes to hundreds of gigabytes.

To accommodate growing data volumes and new file types, organizations must scale storage capacity. But simply expanding existing storage environments is not always the most expedient or cost-effective option.

In recent years, many institutions leveraged cloud storage for long-term data storage. The cloud model had the benefit of low capital expenditures and was well suited for moderate data volumes that rarely needed to be retrieved.

However, as data volumes have grown, many organizations must now rethink their use of the cloud. They need ways to avoid the substantial costs of retrieving increasingly large data sets from the cloud, and they must support research teams that require stronger performance than cloud services can deliver.

## DATA SILOS

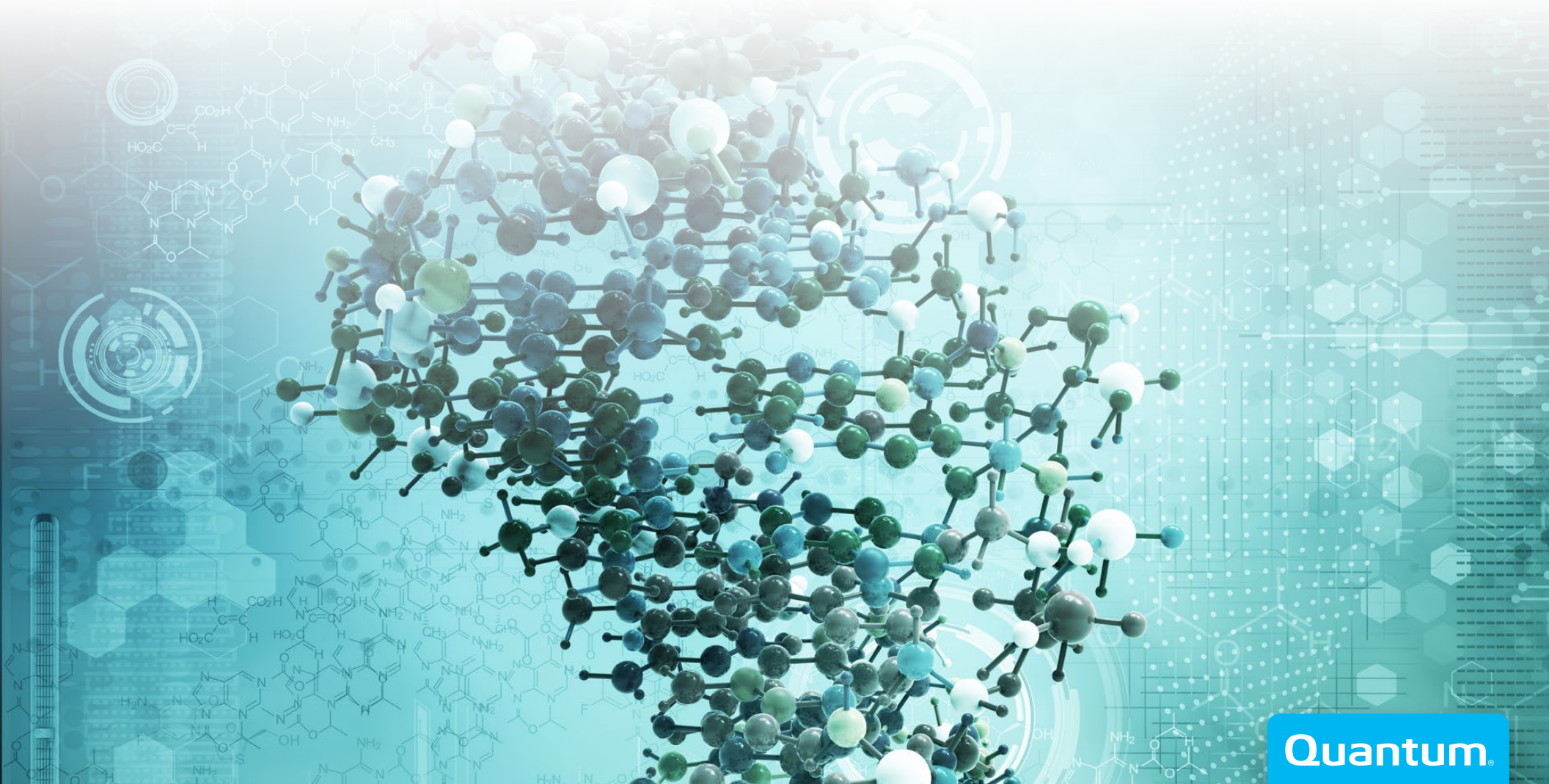
Due to the cyclical nature of research grants and the resulting fluctuation in budgetary planning for IT, many organizations have expanded their technology resources without an overarching, standardized methodology. As a result, they are now faced with multiple data silos. These distinct, unconnected environments make it difficult for researchers to easily access the data they need, when they need it.

## INCREASING COSTS

Rapid data growth and the development of multiple data silos can contribute to spiraling IT costs. Without a single, coherent storage strategy, many organizations buy more than they need. They might expand storage for multiple siloed environments, leaving capacity underutilized in each. Or they might spend too much on production storage because they lack an integrated archiving solution. Managing these complex, siloed environments can be costly and pull researchers away from their primary tasks.

DEFINING

# REQUIREMENTS FOR LIFE SCIENCE STORAGE



Quantum

# Solutions must offer cost-effective scalability and robust performance while simplifying management.

What does your organization need in a storage solution? The right one will provide scalability, performance, and flexibility—while also reducing management complexity.

## SCALABLE CAPACITY

Your storage solution must have the scalability to accommodate rapidly rising data volumes generated by the latest scientific instruments. Research teams need to store both data produced through scientific analysis and raw data, so they have the option to run additional analyses in the future. You should be able to expand the capacity of your storage environment to petabytes of data without ripping and replacing your solution or undergoing major augmentation projects.

An integrated archive solution is critical for preserving and protecting data. In many cases, researchers need to retain data for several years. Over that time, data must stay both secure and unaltered to conform to regulations that mandate the stored copy is identical to the original. Lost or corrupted data could be catastrophic—for clinical research there might be no way to re-capture historical patient samples.

Meanwhile, archived files must be quickly and easily accessible. Researchers can't wait hours or days to retrieve data sets.

The archive must also be cost-effective. In the past, organizations used the cloud as a cost-effective option for archiving. But retrieving data from the cloud can be time-consuming and costly: it can take up to a day to transfer 10 TB at 1 Gbps.

For many life science organizations, on-premises data tape-based archives are a better solution. Compared with the cloud, tape can offer a more cost-effective option for storing and retrieving massive data volumes. Keeping an offline copy of data can also protect organizations from various cybersecurity threats.

## PERFORMANCE

Your storage solution requires robust performance for multiple phases of the life science workflow. For example, you need sufficient sequential performance for ingesting large amounts of data from scientific instruments. In addition, you need strong random I/O performance to support the sophisticated analyses you conduct using HPC or supercomputing resources. The storage solution must be able to complete a significant number of reads and writes per second, especially when you are using large clusters, with numerous cores to process data.

## FLEXIBILITY

Whether you have a small team of researchers in a single lab or numerous researchers spread across the globe, your storage solution must support seamless collaboration. It must help eliminate data silos, enabling researchers to easily access data from a single, shared pool. The storage solution must also have the flexibility to accommodate multiple types of client systems, from Linux and UNIX machines to PCs and Macs, so researchers can tap into the data they need without having to change systems or alter workflows.

Meanwhile, the storage platform should accommodate different types of connectivity. You should be able to continue utilizing your existing investments and have the flexibility to support new systems, regardless of whether those systems are connected by Ethernet or Fibre Channel.

## SIMPLIFIED MANAGEMENT

Few life science organizations have large, dedicated IT teams. In many cases, researchers with technical knowledge are responsible for managing the IT systems. As your organization expands its storage environment to accommodate growing data volumes and new techniques, you need ways to manage that environment simply and in a holistic way. The less time and resources your researchers spend on IT administration, the more they can devote to scientific work and research projects.

# OPTIMIZING YOUR STORAGE FOR LIFE SCIENCES

## WITH QUANTUM

Quantum

# Quantum Xcellis Scale-out NAS offers the scalable capacity, strong performance, flexibility, and simplified management required for life science workflows.

The Quantum Xcellis® Scale-out NAS (network-attached storage) solution can help your organization address key life science storage challenges. Xcellis combines scalable capacity, robust performance and flexibility to help you make the most of new lab instrumentation while reducing administrative burdens.

## SCALABLE CAPACITY

Xcellis enables you to easily and cost-effectively scale capacity, from the ingest system to the archive, to support rising data volumes. With Xcellis, you can start small and scale to billions of files and petabytes of data. And importantly, Xcellis lets you scale capacity independently from performance—you can add drives to arrays and arrays to existing systems without costly, large-scale upgrades. As a result, you gain the agility to quickly integrate new instruments, capitalize on new research grants and expand your research teams.

The Quantum StorNext® file system and data management platform lets you incorporate Xcellis Scale-out NAS as part of a multi-tier environment so you can create an end-to-end solution for managing petabytes of data with a simplified, high-performance workflow. StorNext supports a full range of archival storage solutions, including tape, object storage and cloud. Plus, StorNext scales to the multi-petabyte level, enabling you to preserve years of data and retain fast access for researchers.

## ROBUST PERFORMANCE

Xcellis leverages the StorNext shared file system to deliver robust performance that easily outstrips many competitive NAS offerings. You can rapidly ingest large data volumes generated by lab instruments and then support fast, low-latency access to data for complex analyses. Xcellis supports up to 12 GB per second per client for native client Fibre Channel connections and up to 1 million IOPS per node. That throughput accelerates genome alignment, variant calling, and other life science workloads so you can meet tight research timelines and introduce new discoveries faster.

## FLEXIBILITY

The StorNext platform helps eliminate silos by providing unified access to data through a single global namespace that spans all tiers of storage. Numerous researchers and applications can access data no matter where they are located. Researchers retain access to files even when they are archived; files continue to appear in the file system. They are automatically retrieved from the archive when a researcher opens the file.

StorNext also supports a full breadth of client operating systems. Researchers can access data regardless of whether they are running Linux, UNIX, Windows or Mac operating systems.

Xcellis adds flexibility for multiple connectivity options, including high-performance Ethernet and Fibre Channel, to meet a range of performance and configuration requirements. You can continue to use existing infrastructure, regardless of connectivity, and incorporate new systems easily.

## SIMPLIFIED MANAGEMENT

Even the small, lean IT groups common in life sciences require enterprise-class management capabilities. With Xcellis and StorNext, you gain powerful and intelligent data management functionality while simplifying administration.

For example, the StorNext Storage Manager HSM offers automated, intelligent tiering that enables you to eliminate manual data movement tasks among storage tiers. You can optimize storage utilization and keep data easily accessible without excessive administrative effort.

Storage Manager also provides policy-based, automated data protection, disaster recovery and archiving functionality. Data protection and disaster recovery capabilities enable you to create automatic copies of data, giving your team the peace of mind that valuable research data is safe and secure. You can also set policies to automatically archive data based on the last access date, folder location, or other attributes.

Xcellis and StorNext let you manage all data from a single, straightforward, graphical user interface. You can create clusters, monitor the health of nodes, manage and mount shares, and oversee permissions within a single system view. Storage administrators also have the ability to receive software updates immediately as they become available through the same management utility.

## RETRIEVE PROJECT-BASED DATA SETS FAST WITH STORNEXT AND CLARITYNOW!

As your storage environment grows, grouping data by project can significantly reduce the time to find and retrieve data. Pairing the StorNext file management system with ClarityNow! from DataFrameworks allows you to organize data and apply more useful metadata tags. As a result, projects can be archived and retrieved in their entirety, even as data is migrated among storage tiers and heterogeneous storage solutions. You can also gain better visibility into research storage costs by leveraging showback and chargeback capabilities.

**Learn more about ClarityNow!:**
www.quantum.com/claritynow

DataFrameworks

# ADDRESSING REAL-WORLD LIFE SCIENCE STORAGE CHALLENGES

## WITH QUANTUM

Quantum

# SIB Swiss Institute of Bioinformatics

## Delivering robust resources to drive bioinformatics breakthroughs.

Background: The SIB Swiss Institute of Bioinformatics is an independent, non-profit foundation that includes 70 research and service groups with 800 scientists. The organization is helping to shape the future of life sciences by providing scientists and clinicians with world-class bioinformatics resources and services.

Challenge: SIB needed storage with the cost-effective scalability to support six sequencing centers and approximately 300 research teams who generate up to 30 TB of data per week.

Solution:
Quantum StorNext Scale-out Storage with Quantum tape archives

Benefits:

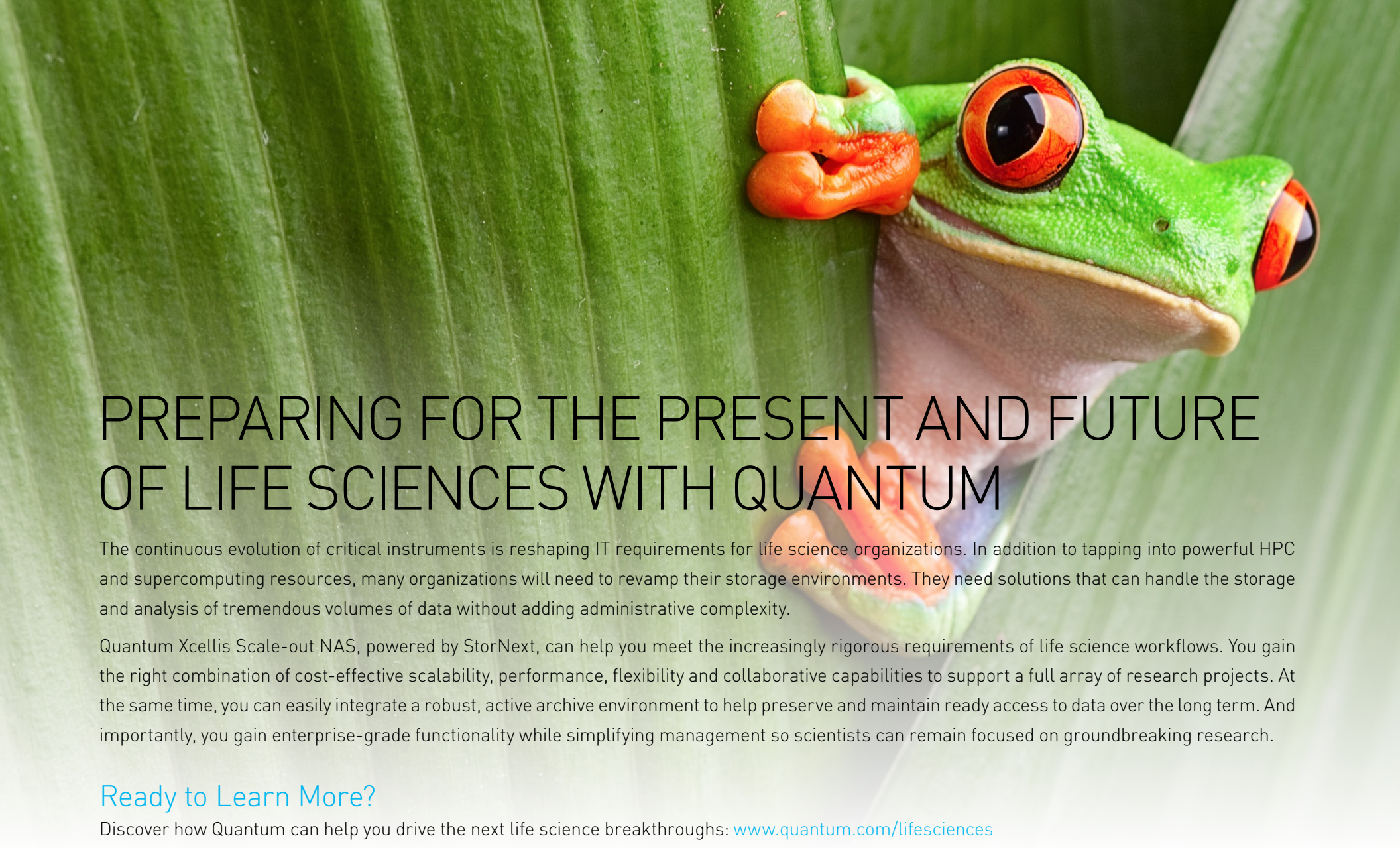
- **Gained the scalable performance and capacity** to keep pace with rapid data growth while controlling costs
- **Increased storage performance** for high-speed sequencing and analysis
- **Implemented archive storage tiers** that keep research data ready for re-use
- **Enhanced protection** of valuable genomics information by automating copies of archived data
- **Added the flexibility** for incorporating next-generation object and cloud storage

**READ MORE AND WATCH THE VIDEO:**
www.quantum.com/SIB

SIB
Swiss Institute of Bioinformatics

> If you provide researchers with the right set of tools, they push the envelope. StorNext tiered storage helps us take data in fast, quickly move it to archive, and keep it ready so bioinformaticians can continue their work.

**Professor Ioannis Xenarios**
Director Vital-IT Group, SIB Swiss Institute of Bioinformatics



# PREPARING FOR THE PRESENT AND FUTURE OF LIFE SCIENCES WITH QUANTUM

The continuous evolution of critical instruments is reshaping IT requirements for life science organizations. In addition to tapping into powerful HPC and supercomputing resources, many organizations will need to revamp their storage environments. They need solutions that can handle the storage and analysis of tremendous volumes of data without adding administrative complexity.

Quantum Xcellis Scale-out NAS, powered by StorNext, can help you meet the increasingly rigorous requirements of life science workflows. You gain the right combination of cost-effective scalability, performance, flexibility and collaborative capabilities to support a full array of research projects. At the same time, you can easily integrate a robust, active archive environment to help preserve and maintain ready access to data over the long term. And importantly, you gain enterprise-grade functionality while simplifying management so scientists can remain focused on groundbreaking research.

## Ready to Learn More?

Discover how Quantum can help you drive the next life science breakthroughs: www.quantum.com/lifesciences

**ABOUT QUANTUM**

Quantum is a leading expert in scale-out tiered storage, archive and data protection. The company's StorNext® platform powers modern high-performance workflows, enabling seamless, real-time collaboration and keeping content readily accessible for future use and re-monetization. More than 100,000 customers have trusted Quantum to address their most demanding content workflow needs, including top studios, major broadcasters and cutting-edge content creators. With Quantum, customers have the end-to-end storage platform they need to manage assets from ingest through finishing and into delivery and long-term preservation. See how at **www.quantum.com/customerstories**.



Quantum®

www.quantum.com
800-677-6268

ST02236A-v01 June 2018