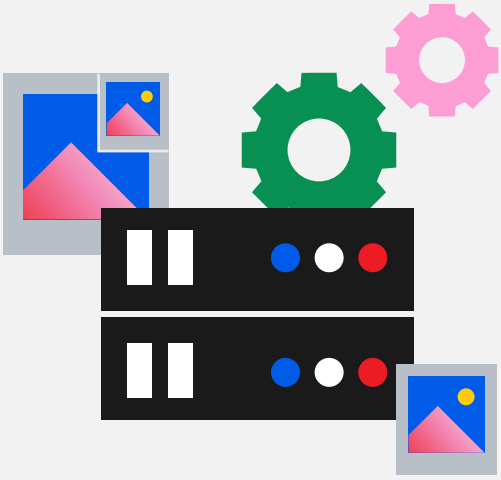
PANZURA

SOLUTION BRIEF

Panzura Symphony & GRAU DATA

Optimizing Unstructured Data Storage and Access in Life Sciences.



The Challenge of Unstructured Data in Life Sciences

The life sciences industry generates vast amounts of unstructured data from genomic sequencing, high-resolution imaging, electronic health records, and laboratory research. The exponential growth in file capacity and file count present multiple challenges, particularly around storage, accessibility, regulatory compliance and cost efficiency. Traditional NAS (Network-Attached Storage) devices, while widely used, come with a significant cost. As research datasets and files increase in size, the cost of high-performance NAS systems becomes prohibitive.

Another major struggle is insight accessibility. Research teams often need rapid access to the insights of the digital artifact (or full dataset) for analysis, but not necessarily the artifact itself. Traditional NAS solutions can provide the artifact, but it's the metadata which holds the insights. The problem of escalating costs is compounded by compliance and regulatory requirements that mandate long-term data retention and retrieval capabilities. Without an efficient data management strategy, organizations face ballooning costs and operational inefficiencies.

Categorizing and Managing Data with Metadata

One way to address these challenges is by implementing a metadata catalog such as GRAU DATA MetadataHub. MetadataHub provides a structured way to transform unstructured data, from NAS or object storage, into actionable insights through advanced metadata extraction of "rich" or "embedded" data (supporting over 500 file types) and analysis, allowing researchers to efficiently locate and retrieve relevant datasets. By tagging data with descriptive metadata, organizations can streamline their search capabilities, eliminating the time-consuming process of manually sifting through vast amounts of information.

Furthermore, categorizing data with custom metadata enables intelligent decision-making, such as:
1. Improved Data Discovery – Categorizing research data based on parameters such as experiment type, specimen source, genomic markers, or drug interactions allows scientists to quickly locate relevant datasets without sifting through the original digital artifacts, which tend to be quite large.

2. Enhanced Collaboration – Proper classification of datasets ensures that researchers across different institutions or disciplines can access, interpret, and share data seamlessly, reducing duplication of efforts and fostering interdisciplinary discoveries.

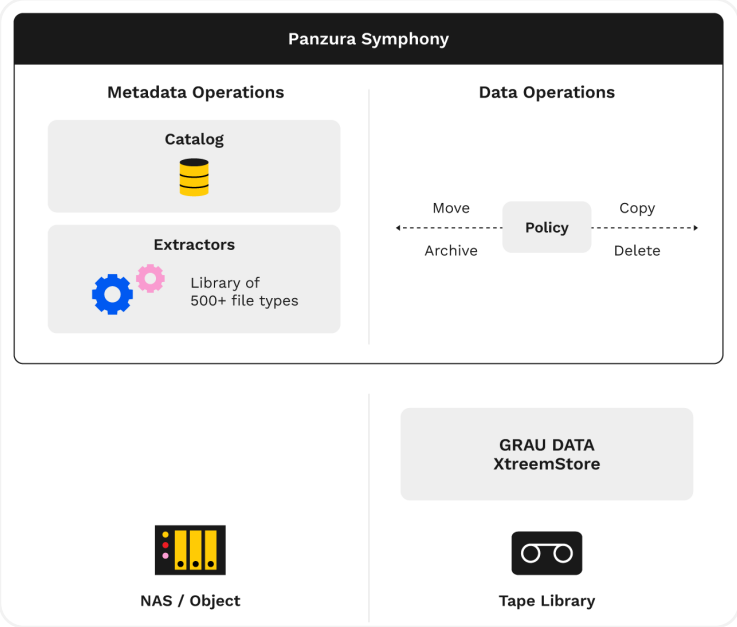
3. Accelerated Analysis & AI Integration – Categorized data enables more effective use of machine learning and AI models for pattern recognition, predictive analytics, and drug discovery by providing structured inputs that improve algorithm performance. Panzura has released another solution brief on using Symphony as a Zero Trust Principled Data Broker, which can be found here.

4. Regulatory & Compliance Readiness – In regulated environments, such as pharmaceutical research or clinical trials, categorization ensures that data is stored, accessed, and archived in compliance with regulatory requirements like HIPAA, FDA 21 CFR Part 11, or GDPR.

5. Efficient Storage & Retrieval – Categorization enables tiered storage strategies, ensuring frequently accessed data remains in high-performance storage, while archived data is efficiently indexed and retrievable when needed. We will focus on this one below.

Cost-Effective Storage

An effective alternative to traditional NAS storage is the use of tape-based storage solutions, with an S3 front-end such as GRAU DATA's XtreemStore, which provides instant data access for "off-line" data.



Tape storage is significantly more cost-effective compared to high-performance disk-based solutions, making it an ideal option for archiving large datasets while maintaining accessibility.

Transitioning data from traditional NAS to XtreemStore requires an efficient data migration strategy.

A data mover is necessary to facilitate this transition while ensuring that migrated data remains accessible within the existing NAS environment. Without a streamlined migration solution, organizations risk disrupting ongoing research workflows and data access.

This is where Panzura Symphony plays a crucial role. Symphony provides automated, policy-driven data mobility, offering seamless data migration, management, and audit tracking for risk and compliance. Integrated with GRAU DATA MetadataHub, Symphony enables automated tiering based on metadata. By flagging artifacts for storage tiering, organizations can move copious amounts of data to lower-cost storage solutions while maintaining its visibility and accessibility within the NAS filesystem.

Enhancing Performance and Accelerating Research

By integrating MetadataHub, XtreemStore, and Symphony, life sciences organizations can establish a holistic data management strategy that reduces costs, enhances performance, and improves data accessibility. This approach offers several key benefits:

1. Cost Savings: Moving infrequently accessed data to tape storage significantly reduces storage costs while maintaining access via your traditional file system.

2. Improved Performance: Keeping only active datasets on high-performance NAS ensures faster processing and retrieval times.

3. Enhanced Data Access: Metadata-driven search capabilities allow researchers to quickly locate and retrieve relevant data, without extracting the full artifact.

4. Regulatory Compliance: Efficient data categorization and archiving help organizations meet compliance requirements for long-term data retention. And metadata tagging allows business units and SMEs to collaborate using the catalog to set and find the relevant data if an inquiry should occur.

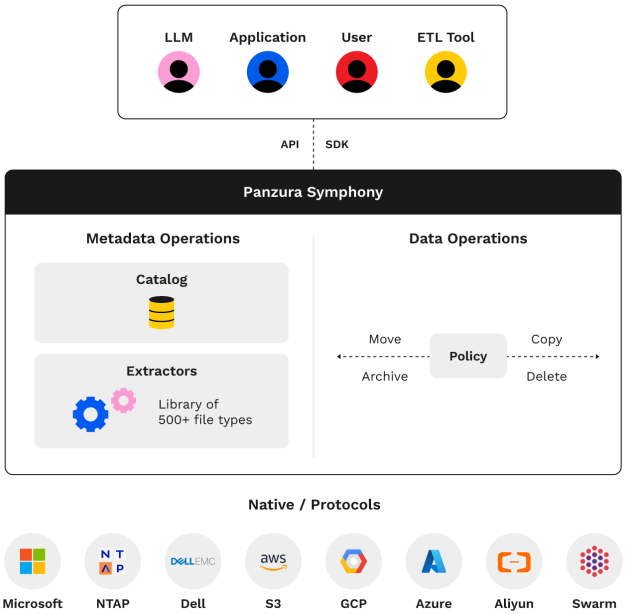
5. Accelerated Research & Development: Optimized data storage and retrieval processes reduce delays in research workflows, enabling faster insights and discoveries.

Conclusion

The life sciences industry is at a critical juncture where data storage and management challenges must be addressed to keep pace with scientific advancements.

Traditional NAS solutions are no longer sustainable due to their high costs for massive data sets. By implementing a metadata-driven approach with MetadataHub, leveraging cost-effective storage solutions like XtreemStore front-ending a tape library, and automating data tiering with Symphony, organizations can achieve a streamlined, high-performance data management strategy.

This integrated approach ensures that research teams have quick and efficient access to critical data while significantly reducing storage costs—ultimately accelerating scientific breakthroughs and improving operational efficiency in life sciences.



Panzura empowers today's digital-first organizations to do impossible things with file data, making them more agile, efficient, and productive. They trust Panzura to help them consolidate dispersed data, see and manage data in and out of the cloud, make it more cyber-resilient and AI-ready, and ensure it is available to people and processes where and when it's needed.

Discover how Panzura can fuel your success at **panzura.com.**