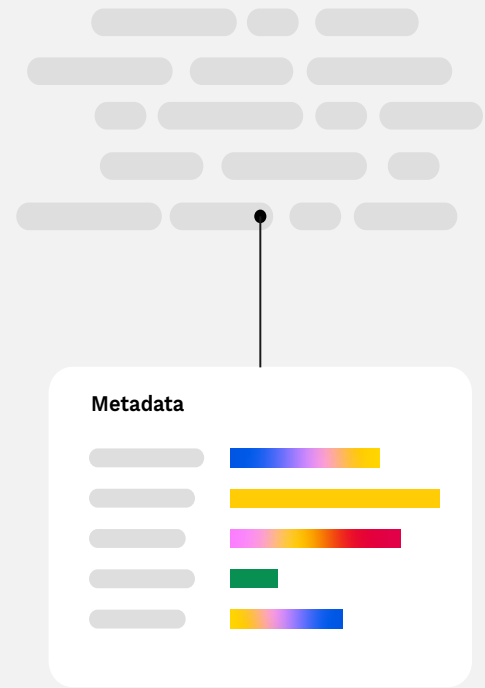


# Panzura Symphony Transforms On-Demand Data Provisioning with GRAU DATA MetadataHub Integration

Managing Dataflow with Intelligent Metadata  
Extraction for Improved Data Quality,  
Greater Efficiency, and Enhanced Security in  
Research-Intensive Industries



IT leaders, data stewards, and data-driven teams such as DevOps and AIOps are constantly battling the sheer volume and complexity of data. The concept of On-Demand Data Provisioning—the immediate, intelligent transfer of data—presents obstacles that these professionals navigate daily. Traditional data management systems often fall short in addressing the dynamic data needs of modern enterprises.

One of the primary challenges faced by technologists is the inflexibility of traditional storage environments. They are typically optimized to handle structured data, but in reality, the vast majority of data is unstructured. Documents, experimental research, correlational data, multimedia files, and other types of discrete or continuous data are not easily managed through conventional systems.

This leads to performance bottlenecks as the systems struggle to index, search, and retrieve relevant data quickly. The lack of efficient data retrieval mechanisms means that IT teams spend an inordinate amount of time and resources trying to locate and move the required data without unnecessary transfer of large datasets.

Data stewards, who are responsible for ensuring data quality and governance, face their own set of challenges. Current solutions often lack the granularity needed to manage dataflow effectively. Metadata, which provides context and meaning to data, is crucial for On-Demand Data Provisioning.

Metadata Type	Example Properties	Primary Uses
<b>Descriptive</b> info that describes what the materials are and the context of their creation	Title Author Subject Publication date Genre	Discovery Display Interoperability
<b>Structural</b> info about where an object is located in a sequence, hierarchy or file structure	Sequence Hierarchy	Navigation
<b>Technical</b> info that describes what the materials are and the context of their creation	File type File size Creation date/time Compression scheme	Interoperability Digital object management Preservation
<b>Rights</b> info about an object's copyright status, holder, & any relevant licenses	Rights holder Copyright status License terms	Interoperability Digital object management
<b>Preservation</b> info about the preservation status of a digital file	Checksum Preservation metadata	Interoperability Digital object management Preservation

However, traditional systems do not offer sufficient metadata management capabilities, making it difficult to maintain data integrity and compliance. The inability to efficiently extract, expose, and leverage metadata results in silos where valuable information is trapped in isolated systems, further complicating the data management lifecycle.

Teams tasked with data movement and orchestration encounter significant hurdles due to the limitations of existing storage options. These environments are not designed for the dynamic nature of data which needs to be accessed and analyzed in real time.

The static nature of traditional storage solutions means that technologists and data teams must often resort to manual processes to move and deliver data, leading to delays and increased operational costs. The lack of automation and intelligent data management tools exacerbates the problem, making it difficult to harness data in support of business outcomes.

The emerging data landscape is poised to make On-Demand Data Provisioning even more challenging in the future. The exponential growth of data, driven by technological advancements and the proliferation of edge devices, means that organizations will need to manage and move even larger volumes of data.

The increasing complexity of data, with diverse formats and sources, will require more sophisticated data management strategies. The emphasis on data privacy and security continues to necessitate ever-more-stringent and complicated compliance measures.

## The Power of Integration: Symphony and MetadataHub

Panzura Symphony offers an integration with GRAU DATA's MetadataHub that delivers a novel solution for On-Demand Data Provisioning. The integration provides the ability to access and analyze only the necessary data artifacts using datatype-specific metadata, obviating the transfer or reacquisition of entire datasets.



Symphony is a data services platform designed for exabyte-scale discovery and assessment, risk and compliance analysis, and data movement orchestration. MetadataHub is a solution from GRAU DATA that unlocks the potential of unstructured data by extracting and exposing metadata fields from over 400 file formats.

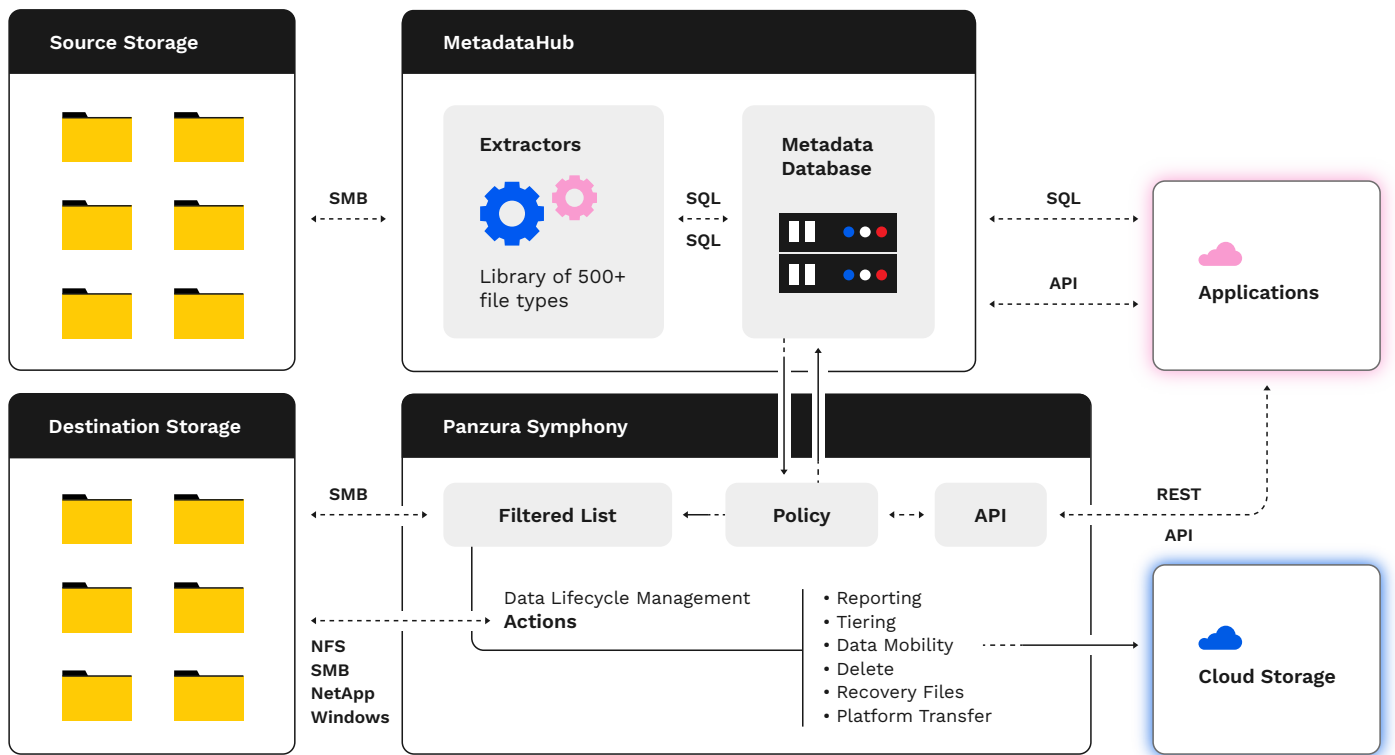
GRAU DATA's MetadataHub analyzes and extracts valuable information from file metadata, creating a comprehensive metadata catalog. Panzura Symphony leverages this metadata catalog, acquiring only the needed data attributes and sidestepping the unnecessary transfer of large datasets. This significantly reduces network traffic and storage demands which dramatically improves performance and efficiency.

The integration of these two solutions allows data delivery and assessment based on specific file attributes, application-specific metadata, datatypes, and business context classifications, ensuring data is categorized and handled according to its characteristics.

Dynamic Workload Placement and metadata-aware capabilities of the integrated solution reduce operational costs and enhance data visibility and control.

At its core, Panzura Symphony streamlines the process of accessing, analyzing, and sharing data. It acts as a single portal where all reporting and data can be connected. Among of the standout features of Symphony is its ability to integrate with existing analytics tools using common interfaces and open standards.

The technical specifications of Symphony are impressive. It supports a wide range of data sources including file systems, protocols, and object stores. Panzura Symphony is designed for scalability, allowing it to handle large volumes of data without compromising performance.



While Symphony excels at unified data discovery, assessment, and orchestration, GRAU DATA's MetadataHub focuses on unlocking the value of unstructured data. MetadataHub extracts and exposes metadata from these files, turning embedded metadata into valuable insights. Its capabilities are extensive. It allows users to search, analyze, and comprehensively evaluate files without having to read the file itself.

This functionality is particularly useful for organizations dealing with large volumes of data, as it avoids unnecessary data acquisition and eliminates inefficient storage. MetadataHub's container architecture ensures massive scalability, making it suitable for organizations of all sizes.

One of the key strengths of GRAU DATA's MetadataHub is its ability to automate data extraction and analysis processes, reducing the need for manual intervention. MetadataHub also supports custom development of special extractors, enabling technologists to tailor the platform to specific contextual needs.

The combination of Panzura Symphony and GRAU DATA's MetadataHub creates a synergy, and the benefits are not just theoretical. According to a report by McKinsey, embedding digital collaboration into process workflows can enable faster, better decisions that improve key performance indicators (KPIs) and drive bottom-line value. This is attributed to deployment of the types of capabilities offered by the integration between Symphony and MetadataHub.

## How it Works

Intelligence resides in metadata which can contain thousands of tags that capture critical content and context from data files. However, metadata is much lighter compared to the actual files themselves, making this approach extremely fast and efficient, as it allows for quick access and processing of data without the overhead of handling large files.

Seamlessly capturing critical content and context from data files, MetadataHub creates a rich metadata repository that serves as a “proxy” for the original files, providing enhanced visibility in Symphony. Metadata is a fraction of the size of the file itself, so this proxy significantly reduces the need to transfer large files, optimizing network and storage resources.

## Metadata vs. Data

Patient ID	Date	Procedure	Clinical Notes	Assessment	
112	3/22/2024	70551	11488-4	Yes	Metadata
113	4/14/2024	70552	18748-4	Yes	
114	5/10/2024	72141	11488-4	Yes	Data
115	6/29/2024	72156	18748-4	Yes	

The Panzura Symphony integration with GRAU DATA’s MetadataHub also delivers comprehensive data visibility and insights that accelerate informed decision-making across departments, teams, and roles like data stewards and those responsible for compliance and security policies. For example, it enables on-demand access to data across an entire organization so storage operations, business analysts, data governance officers, and even AIOps can access high-quality data for more accurate analysis or to feed automated processes and pipelines.

- 01 Metadata-Driven Data Access:** MetadataHub analyzes and extracts valuable information from files, creating a comprehensive metadata catalog.
- 02 Intelligent Data Acquisition:** Panzura Symphony connects to the catalog and allows custom ‘sources’ to be created using file-type-specific metadata filters.
- 03 Optimized Data Transfer:** Dataset orchestration can be controlled using comprehensive metadata filters as well as standard file system and object attributes, avoiding the need to transfer and open files to understand the metadata within.
- 04 Secured and Controlled Access:** Existing namespace support ensures that only authorized users can access specific data segments.

## Refine and Streamline Workflow Automation

Automated workflows offer numerous benefits for data processing in data-intensive industries such as life sciences. Accelerating data processing and delivery enables quicker insights and faster responses to research needs. This means researchers can analyze experimental results, identify potential drug targets, and develop new therapies more efficiently. Moreover, automating manual tasks liberates data professionals to focus on higher-value activities, such as data analysis and interpretation, rather than tedious, repetitive activities. Instead of spending hours on data cleaning and formatting, they can focus on uncovering hidden patterns, developing predictive models, and generating actionable insights from the data. This shift in focus leads to more strategic use of data and ultimately better business outcomes.

Additionally, automation significantly minimizes the risk of human error, leading to more accurate and reliable data, which is crucial for making informed business decisions. Manual data entry and processing are prone to errors, which can have significant consequences for businesses, from financial miscalculations to flawed strategic decisions. Automation dramatically reduces these risks, ensuring that data is consistent, accurate, and trustworthy.

Symphony's interoperability helps ensure visibility and seamless integration between file system deployments and diverse storage landscapes including on-premises, and private, public, hybrid, and multi-cloud object stores. It supports multiple storage tiers and offers enhanced control over data assets. With Dynamic Workload Placement capabilities, IT teams are able to precisely orchestrate data placement, archive, and migration to transform the data operations framework.

The integration of Symphony with MetadataHub empowers DevOps teams by automating key data management processes. With Dynamic Workload Placement, which intelligently and automatically positions data where it is most needed, Symphony ensures that data is efficiently processed for artificial intelligence (AI) pipelines, leveraging various triggers and transformation functions—such as webhooks, and pre- and post-run actions – to handle different data workloads and ecosystems.

Moreover, the automated generation and distribution of reports provides stakeholders with timely insights into system health and performance, facilitating faster incident response and improved collaboration. Instead of manually compiling reports, DevOps teams can configure automated reports that are delivered to relevant stakeholders at predefined intervals or triggered by specific events. This ensures that everyone is informed and aligned, fostering a culture of collaboration and proactive problem solving.

Crucially, this integration also strengthens security and compliance by automating data access, dataset location, and retention policies, ensuring adherence to industry regulations and organizational standards.

## Accelerate Data Preparation

MetadataHub's catalog plays a crucial role in both data discovery and data quality. Providing granular, datatype-specific visibility into files, including details like data source, instrument, experiment, file type, and creation date, it allows users to quickly identify relevant data for analysis in tools like Amazon Redshift, Snowflake, or Databricks. Furthermore, Symphony uses MetadataHub's metadata catalog to facilitate data profiling and validation, revealing potential data quality issues such as missing values, inconsistencies, and outliers. Automated data profiling helps identify data anomalies that might otherwise go unnoticed, ensuring that data is accurate and reliable before it is in the pipeline and used for analysis.

Metadata can be used to enforce standard, consistent data definitions and formats across different sources. This is crucial for organizations that deal with data from various systems and departments, where inconsistencies can easily arise. Enforcing standardized definitions ensures that everyone is working with the same understanding of the data, reducing confusion and errors.

This proactive approach enables users to address these problems during the data preparation stage, ensuring the accuracy and reliability of their analysis from the start. Identifying and resolving data quality issues early on, organizations can avoid costly mistakes and ensure that their data-driven decisions are based on solid foundations.

## Granular Metadata Drives Quality

MetadataHub's catalog acts like a detailed map of target data, providing valuable information about its structure, content, and relationships. This "map" can significantly improve data quality and reduce preparation time when preparing and moving data to analytics applications. Granular metadata allows for precise data selection and transformation, reducing errors and inconsistencies that can arise when automating data pipelines.

MetadataHub's catalog provides fine granularity and high selectivity without moving entire files which can cause network congestion and spiraling cloud costs. For example, if a data scientist needs to extract specific attributes from a dataset for machine learning, they can leverage metadata and identify the relevant data fields, ensuring that the correct artifact is acquired and ready for analysis.

By providing a comprehensive and easily accessible view of the data landscape, the integration between Symphony and MetadataHub empowers data professionals to work with data more cost-efficiently and effectively, ultimately leading to better data quality, faster insights, and more informed business decisions.

# Benefits



## Reduced Storage Costs

Providing access to only the required portions of datasets, organizations can dramatically reduce high performance storage footprints and eliminate the need for multiple tools and interfaces.



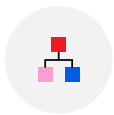
## Improved Management and Performance

Eliminates the need to transfer entire datasets, optimizing network performance and streamlining data access for people and processes.



## Enhanced Data Security

Limits users and processes to specific datasets, reducing the risk of unauthorized access and breaches, and automatic policy enforcement strengthens compliance.



## Data-driven Decision-making

Metadata extraction for richer insights into file context to identify relevant information and automatically categorize data.



## Improved Data Quality and Reliability

Enhances the quality of data by supporting AI enrichment with generated context and information, making it more usable and reliable for AI and ML models.



## Accelerated Data Discovery

Empowers granular metadata-based search and retrieval across the entire data landscape, dramatically reducing the time and effort required to locate information.



## Efficient Collaboration

Breaks down barriers to efficiency and productivity, making data more accessible and usable so teams can collaborate regardless of location.



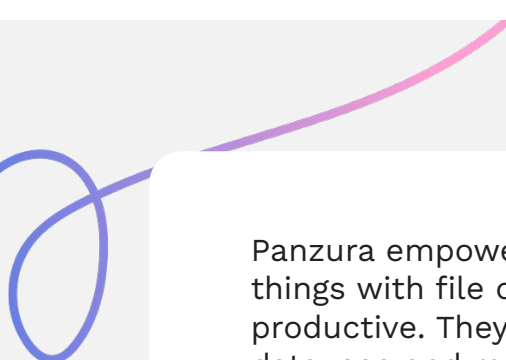
## Implementation and Support

Designed to seamlessly integrate with existing infrastructure, Symphony and its integration with MetadataHub ensures minimal disruption to data operations. Starting with a comprehensive assessment of current data landscape and specific requirements, our team can then develop a tailored implementation strategy that addresses every customer's unique needs and goals.

The phased deployment process allows for a smooth and controlled rollout. Our team works closely with customers throughout the entire process, providing guidance and support every step of the way. To help users to effectively employ the solution, we provide comprehensive training and documentation. This ensures that customers can confidently manage and optimize the solution from day one.

Our commitment to every customer's success extends beyond the initial implementation. We offer ongoing support and optimization services to ensure the solution continues to meet evolving needs. As businesses continually grow and change, we'll be there to assist customers as they adapt, maximize, and reap the full value of their investments.

The integration of GRAU DATA's MetadataHub with Panzura Symphony marks a significant advancement in how organizations manage and utilize their data. This powerful combination creates a dynamic synergy, infusing AI-driven intelligence into data stores while ensuring seamless global access. Enriching data with comprehensive metadata and providing a unified view across the entire data landscape, this integrated solution helps teams make faster, more informed decisions.



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](https://panzura.com).

