**THE LIFE SCIENCE DATA CHALLENGE**

Life science organizations are confronted with processing, moving, storing and keeping track of ever increasing amounts of genomics data, securely sharing results with researchers around the world, safely and economically storing data long term, and easily finding it when needed. Lacking the availability of a capable data management platform, data scientists and bio-informaticians have been forced to resort to time consuming and expensive manual data management activities that don’t scale.

**MEDIAFLUX TO THE RESCUE**

Mediaflux enables scientists to focus on doing science instead of tedious data management. Mediaflux is a robust, flexible data + metadata management platform that unifies life science data silos into a secure distributed collaboration environment across a single global namespace. Designed for extreme scale and performance, Mediaflux automatically indexes content and metadata to enable global management of genomics data on any storage system across multiple locations.

### HIGHLIGHTS

- Accelerate secondary analysis and reduce storage costs
- Automatically tier and move data
- Securely access data worldwide within a single global namespace
- Tag data with instrument and project metadata to search, report and act on billions of files
- Automate analysis workflows

### THE GENOMICS DATA LIFECYCLE

**Primary Analysis** (within the instrument)
- High Throughput Sequencers
- FASTQ and BCL Files
- Tag files with instrument and project metadata

**Secondary Analysis**
- Alignment and variant calling
- HPC Cluster
- High Speed Flash Storage (NVMe Flash)
- Add fast flash storage to accelerate analysis

**Tertiary Analysis**
- Scientific discovery, statistical analysis and database search
- SAM/BAM Files
- Re-run archival genomic data through updated pipelines
- Tier data from fast storage to low-cost, high-capacity storage

**MEDIAFLUX CAPABILITIES**

- Manage sequencing, storage and computational resources, and move raw and derived datasets from resource to resource to complete analysis
- Securely access data worldwide within a single global namespace
- Tag data with instrument and project metadata to search, report and act on billions of files
- Automate analysis workflows
- Accelerate secondary analysis and reduce storage costs
- Automatically tier and move data
- Unify distributed storage silos into one highly visible, manageable global namespace that can be easily and securely searched and accessed by worldwide collaborators
**Cut Storage Costs**
Tier data to low cost storage, deduplication and useless data removal, the right data at the right place at the right cost.

**Ensure Data Provenance and Reproducibility**
Track which genomic data was processed by which application and version within an analysis pipeline to ensure reproducibility and data integrity.

**Find Data Easily**
Metadata extraction and search locates hard-to-find data by querying workflow-specific metadata describing stored data.

**Integrated Data Lifecycle Management and Security**
Data collection, analysis, sharing (open data export), discovery, archive, data re-use and monetization, versioning, access control, auditing, checksum and WORM.

**Complete Data Visibility Across a Single Global Namespace**
Characterize information by data type, use, owner, size, last accessed, etc. Gives administrators the information needed to lower storage costs and preserve research data.

**High Performance and High Scale Operation**
Parallel I/O for ingestion, migration and replication, high-performance, massively scalable XODB NoSQL database.

**Data Protection**
Replication and backup, disaster recovery across remote sites, move and distribute data to wherever it is needed.

**Analysis Workflow Automation**
Automate complex workflows and processes, and deliver data to external applications and systems for transformation.

**About Arcitecta**
Arcitecta is a creative and innovative software company. Founded in 1998, it’s our mission to build the world’s best data management platforms.

Over nearly twenty years of innovation and entrepreneurship, our technology has formed the foundation for managing the simplest, and the most complex data, for individuals through to large teams of people at global enterprises. Our experience in working with our customers to define the questions they need answering has simplified data intensive workflows. We expose pioneering ways for our partners to speed up the pace of discovery and translate that into tangible actions.

Our technology is only one part of our successful and enduring relationships. The best outcomes for demanding customer challenges arise from deep collaboration. We place great value on becoming fully immersed in your problem-space to help craft the best possible, custom-curated biological database for the most demanding research environments.

**We Solve Problems Others Cannot**
We own the entire technology architecture. With deep knowledge of every line of code, we can customize a solution to the most complex data problems.

**Together, We Build Your Future**
After listening to our customers’ needs, we have the product and people-power to create and implement the best possible solutions - we build what you need.