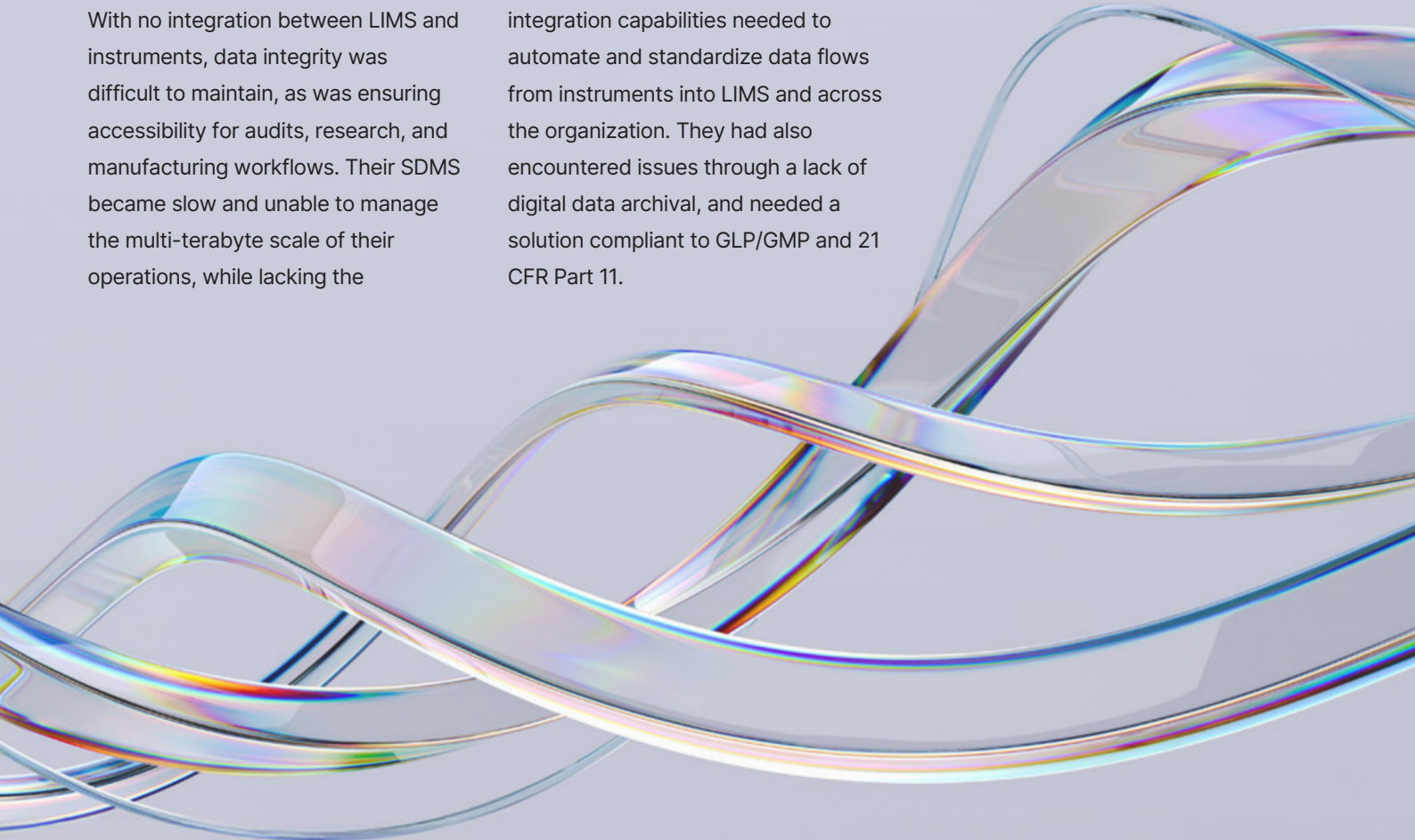


ENABLING DATA INTEGRITY AND VISIBILITY IN GENETIC DRUG DEVELOPMENT

A leading gene therapy innovator faced escalating challenges collecting and managing vast volumes of complex instrument data.

With no integration between LIMS and instruments, data integrity was difficult to maintain, as was ensuring accessibility for audits, research, and manufacturing workflows. Their SDMS became slow and unable to manage the multi-terabyte scale of their operations, while lacking the

integration capabilities needed to automate and standardize data flows from instruments into LIMS and across the organization. They had also encountered issues through a lack of digital data archival, and needed a solution compliant to GLP/GMP and 21 CFR Part 11.



The Splashlake Solution

Splashlake provides a flexible, open, and scalable data platform enabling laboratories to capture, contextualise, store, and access all their laboratory and production data efficiently. It replaces multiple single-function tools, acting as a “Swiss army knife” for lab data management, while avoiding vendor lock-in and proprietary data traps.

“

I don't want a single tool that does one thing. I want a system that can handle instrument integration, digital archiving, and data management, without locking us into proprietary formats

Director of Laboratory Systems

”

The organization required:

- A scalable, cost-effective instrument integration solution
- Data management and storage capable of handling large data sets in a non-proprietary, human-readable format, to support future AI strategies
- The ability to manage digital archiving aligned with GLP/GMP
- A unified system that could achieve all this while maintaining data visibility, accessibility and compliance

The global biotechnology company had run into a number of challenges around collection and storage of their analytical data, particularly large data sets from techniques such as flow cytometry, next-generation sequencing (NGS), Western blot analysis and high-resolution microscopy images.

As part of the Western blot analysis process, QA scientists had to log onto the instrument that was used to run the analysis in order to review the data alongside the audit history. This involved booking time out on an instrument which takes hours to complete a single run, so any time used for data review takes precious time from the instrument's availability for analysis. If they didn't enable the data to be viewed elsewhere, they would have to purchase additional instruments just to allow for QA data review.

The company's SDMS had historically worked well, but the system struggled to cope with around 30 terabytes, which made it impossibly slow to store large data sets, or get to data quickly.

The SDMS also encrypted the data, which meant they were locked in to using and managing additional systems in order to access it. It was also impossible to view instrument data through the SDMS, so scientists would have to access it through another method, and getting the data out in order to view it was painful. The system also couldn't easily meta tag, which meant a reliance on the user to name the file correctly in order to achieve appropriate meta tagging.

Three Key Areas of Splashlake Application

1. Instrument Integration

Splashlake connects the LIMS to key analytical instruments, automating data capture without taking instruments offline for audit review. This frees instrument capacity, supports data integrity and traceability and enables seamless integration with the LIMS and lab workflows.

2. Large Data Storage and Cataloging

Splashlake enables scientists to store and manage large, complex data files while providing version control, metadata tagging, and audit history. This replaces slow, limited systems and overlays control and accessibility on S3 and other storage solutions. Splashlake's impressive metadata tagging enables data cataloging and search capabilities, with scientists also being able to view the data they query.

3. Digital Archiving for GLP/GMP

The organization uses Splashlake to implement secure, compliant digital archiving, addressing patent, legal, and audit-readiness needs while enabling easy, structured retrieval of data.

Splashlake easily converts proprietary formats to human-readable, allowing laboratories to retire legacy systems confidently, knowing data remains protected and accessible.

Splashlake Implementation

Splashlake's flexible deployment options enabled the organization to opt for the solution to be provided on AWS as a Managed Service, using their own AWS S3 bucket to ingest and store instrument and study-relevant data. Splashlake was initially provided as SaaS on AWS as a TEST instance. The system was deployed in their private AWS environment, updated with a VAL and PROD instance, and validated according to the required standards and policies of the organizational QA team.

Splashlake's Bring Your Own Storage approach enabled the organization to choose the storage solution best suited to their needs. They opted to provide a private AWS S3 storage and SQS message queue in their AWS tenant, to ensure full control and ownership of their data.

Splashlake coordinated with the IT and AWS teams to mount the AWS S3 bucket to the Splashlake application on AWS.

The Splashlake S Drive provides a local drive letter that is used for instrument software to save files to its home directory, or to allow scientists to save, manage and reuse and reprocess data sets. The S Drive comes with an audit trail, full versioning, and provides users with the same permissions as in Splashlake. Splashlake installed and configured the S Drive on all computers where instruments were connected, or on scientists' individual computers.

The Splashlake file capture agent monitors folders of instrument software and uploads files in real time on a fixed schedule, delivering high transfer speeds for large files. Splashlake installed and configured the file capture agent on computers with instruments connected.

The organization envisioned letting scientists store, access and manage study-relevant data in Splashlake, and finally archive them after completion using Splashlake's archive module. Splashlake mounted an archiving S3 bucket to the Splashlake solution. An additional S3 bucket was provided for long-term archiving of data, which could be 'hot' (fast) or 'cold' (slow) storage, and terms were defined for the automated archival of study data. They also defined users for the digital

archivist role, which manages the deposit, retrieval and reading room processes, controls permissions and can see all archived content and audit trails.

Next, it was time to integrate the LIMS with Splashlake, and embed the Splashlake viewer into the LIMS. Splashlake delivered documentation to the LIMS team and supported them in testing the data exchange with Splashlake. The team provided the following configuration in the LIMS instance:

- REST call to send samples from LIMS to Splashlake
- REST call from LIMS to Splashlake to pick up results
- REST endpoint to receive notification of new results at Splashlake

They then tested end-to-end data exchange between Splashlake and instruments, and tested the embedded viewer to visualize converted data, including signals and metadata.

Following the LIMS integration, the integration between the automated Western blotting instruments and the LIMS was tested:

- Splashlake receives measurement request from LIMS
- Western blot instrument data captured and uploaded into Splashlake

- CSV and CBZ file formats converted into future-proofed AnIML format
- Relevant information (results, metadata) extracted from file name and AnIML file content
- Extracted information/results integrated into LIMS measurement request, as per customer requirements

The organization also wanted to upload microscopy images to S3 Cloud storage. Splashlake evaluated connectivity options including the Splashlake S Drive and file capture agent to enable fastest cloud storage upload for the high-resolution image files.

The customer then added relevant metadata to the stored images in Splashlake, and configured metadata extraction and tagging to enable easier search, access, and findability of images.

The incumbent SDMS was also replaced by the Splashlake solution. Splashlake enabled the ingestion of original data from a local hard drive on a scheduled routine. The new solution supports version control and provides a full audit trail, as well as making data accessible for review in the Splashlake viewer.

Following a train the trainer principle, Splashlake offered training for administrators and key users, to get them ready to use the system and ensure best acceptance.



Splashlake is a unique, purple unicorn solution that can do the work of four or five separate systems. It's fast, flexible, and lets us manage data the way we want, using one system instead of many.

Director of Laboratory Systems



Benefits of Adopting Splashlake

By adopting Splashlake, this global biotechnology company has:

- Enabled data integrity and visibility across large, complex data sources
- Freed up instrument time, reducing operational costs and delays
- Solved archival challenges to support GLP and 21 CFR Part 11 compliance
- Delivered multiple functions in one platform, simplifying support and lowering costs
- Future-proofed their laboratory data ecosystem for advanced analytics and AI applications
- Maintained control of their data without proprietary constraints

By transferring and transforming all their legacy CDS and ELN data into Splashlake in a human-readable format, the organization was able to discontinue use of the incumbent solutions, freeing them from the additional costs and restrictions of the proprietary format, and enabling them to access the data indefinitely through Splashlake.



One of the challenges with their previous SDMS solution was the transfer of large files, such as flow cytometry data. The transfer into the system was extremely slow and almost unworkable. With Splashlake, the scientists were able to monitor the creation of such large data files and schedule them to be moved into the system during offline hours, avoiding any impact to system performance.

Splashlake also makes more creative and flexible data processing possible through Python notebooks. The staff can apply Python notebooks to any data stored in Splashlake, providing an interactive environment that allows for a more intuitive and flexible workflow for exploring data, prototyping code, and sharing results.

The meta tagging capability in Splashlake enables highly granular data access and improves findability across the vast amount of data held in the system.

Meta tags can be automatically populated from file names or within files, or manually added. Previously, the organization was using a data cataloging system that could tell scientists where certain data could be

found, and who to request access from, but it didn't make it possible to see that data. With Splashlake, they now have the ability to query and instantly view the matching data across all sources, so are able to retire the data cataloging system.

Spectral Compound Library and Data Usability Expansion

The company is looking to expand its use of Splashlake to spectral compound library management, replacing outdated spreadsheets with structured, searchable datasets, supporting collaborative research, and laying the groundwork for advanced analytics and AI-enabled search across datasets.

Ready to Transform Your Lab Data Workflows?

If your organisation is facing similar challenges with fragmented lab systems, data silos, and manual workflows, Splashlake can help you unify, simplify, and future-proof your data ecosystem.

Contact us to explore how Splashlake can support your lab's transformation.

