

Case Study: How Researchers at The Jackson Laboratory Used Climb to Avert a Data Tsunami



The Challenge

Researchers at The Jackson Laboratory had accumulated a total of 1300 spreadsheets and media files containing over 93,000 data points in share folders for one study. Data from this study were still pouring in and the “data puddle” was growing deeper and would grow into a “data tsunami.” File transfer folders were used for delivering data and status reports to collaborators. Quality Control (QC) and error resolution was coming to a crawl. Collaborators needed expert data wranglers to help stem the tide.

“Our lab conducts high throughput lab animal phenotyping and generates enormous amounts of complex data. We have high-profile research contracts with NIH and with leading edge companies. Reliable and accurate data capture, experimental reproducibility, and process control are crucial to our science. We tried a large on premises LIMS, and it failed to meet our need to share data across multiple sites. RockStep Solutions stepped up to the plate with Climb, and their expert data wrangler team worked directly with our phenotyping team to help address our needs. Our lab now runs more efficiently.”

-Gary Churchill, Ph.D., Professor Karl Gunnar Johansson Chair, The Jackson Laboratory.

Dr. Churchill and other investigators at The Jackson Laboratory are increasingly working on very large, high-throughput studies in collaboration with biotech and pharmaceutical companies. Dr. Gary Churchill’s lab needed a way to manage all aspects of data collection and aggregation in one system, ensuring accurate data sharing with a geographically dispersed team of scientists, and they needed a solution that could document the entire lifespan of each animal in the study.

Dr. Churchill recognized they needed to streamline the data collection and delivery process. At the beginning, data wrangling could take weeks. If a discrepancy was found in the data review, they had to look for procedural errors in the data. After a thorough review of the spreadsheets, they needed to determine all the people involved with the suspect data and then email each of those people. Time constraints meant the task would often get pushed out over several weeks, with technicians trying to recall what happened on a particular day when the lab generally conducted five tests with 80 animals per test.

To handle the load, the Churchill lab needed the ability to retrieve study data over the course of multiple years and to track protocol modifications over time. In addition, the Churchill lab uses smart cages, generating massive amounts of data on mouse activity and body metrics. Making sure these data are available to the client, within the context of the study, is crucial to the project.

Solving the Challenge

The Churchill lab investigated several possibilities for managing their data. They had been using a combination of the JAX Colony Management System (JCMS), for animal tracking and some data collection, and spreadsheets. Although JCMS would allow raw data files to be uploaded, it did not support multiple file types such as image and video files.

Additionally, large data sets could not be easily uploaded, thus many data files were stored elsewhere, making it difficult for scientists to do calculations or quality control (QC) reviews on their raw data.

With the goal of replacing JCMS and spreadsheets, Churchill's lab evaluated several other solutions, including one of the established on-premises industry standards for preclinical research. They found this solution to be unintuitive with a steep learning curve, making it difficult to ensure technician compliance. Additionally, because this solution was on-premise, the availability of the data to the client was limited. They would also need to supplement this system with another colony management system to ensure their client had the 360° view of all data associated with the animals.

Dr. Churchill was looking for one solution to manage animal colony data, study design, sample tracking, data collection, and data delivery. They needed a system that could also automate data quality control (QC) and give their client easy access to reviewed data files and associated study and animal data.

Dr. Churchill turned to RockStep Solutions' flagship product, Climb. Because Climb provides a complete system for tracking data from animal birth to study data aggregation, he no longer needed to piece together data from multiple systems and spreadsheets. According to Laura Robinson, project manager, *"We could configure Climb to our specific schedules and laboratory processes"*, and they could do it themselves, without the need for expensive customizations and consultants.

RockStep collaborated with the Churchill team to organize and load two years' worth of legacy data comprising ~93,000 discrete data points across ~1,300 spreadsheets and media files. These data were harmonized in Climb; they are now available in searchable format and can be queried across multiple experiments and studies. There is no longer a need to correlate data from multiple spreadsheets to review test results across an entire study.

Problem Solved

Data from ongoing studies are now entered directly in Climb, or in the case of instrumentation data, uploaded in a searchable data format. Raw data files from smart cages, videos, and images can be directly uploaded with the task, so relevant information is captured in context. As Dr. Churchill put it, *"The primary benefit is in the organization and accessibility of the data (no more spreadsheets)."*

Data review is now easily handled through Climb. Investigators can download the data, make changes, and upload the file into the task details. Climb tracks where the investigator made changes, providing an audit trail that can be accessed throughout the study. Simple database queries and audit logs have shortened the time to diagnose data discrepancies from weeks to hours. For project manager, Laura Robinson, *"Having an interactive database that can do data QC is a huge advantage. My staff loves that they can look at test result differentials over date ranges to look for and fix data quality issues"*.

Climb's application programming interface (API) and developer's portal allows Dr. Churchill's client to develop their own applications to pull Climb data on demand for analysis. In addition to experimental data, the API integrates strain and husbandry data, providing context for the test results. The portal has eliminated the need to transfer spreadsheets of data to their client. Now up and running, Dr. Churchill said *"It seems like everyone is very happy with how the API worked out, and our collaborators can pull whatever data they need, when they need it."*

In conclusion, Churchill now has control of their data, the tsunami has been averted, and they have more time to focus on their research.

"... we are seeing benefits now in terms of scheduling and real time access to test results, the time we'll save retrieving, quality checking, and analyzing data will significantly increase as the studies progress and we accumulate data,"

-Gary Churchill, Ph.D., Professor Karl Gunnar Johansson Chair, The Jackson Laboratory.