



Case Study | Instrument Workflow Automation

# Automating Data Movement from Scientific Instruments at a Leading US Research Hospital

---

# The Starfish Client

Two affiliated research institutes within a major US research hospital. Multiple research groups across the institutes rely on over 30 different laboratory instruments, including mass spectrometers, lipidyzers, calorimeters, chromatographs, and more.

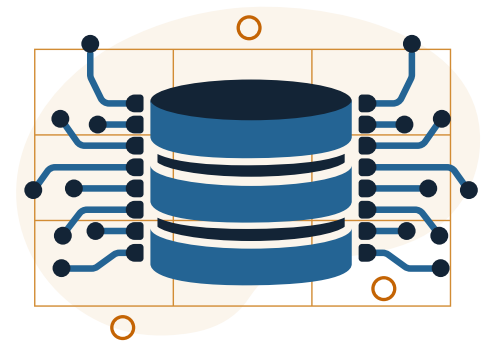


## The Problem

Like many research organizations, the institutes were struggling to manage raw research data from instruments. Researchers were tasked with manually copying files from the local Windows file system on each instrument to their next storage destination, either for further analysis or for long-term storage. The time-consuming process not only distracted researchers from more important tasks, it was also error-prone.

**The institutes needed a way to offload manual tasks from researchers while ensuring files and data for clinical use cases were stored appropriately. They sought a solution that could support the following needs:**

- Enable consistent file naming and storage in centralized locations for easy reference.
- Send copies of all files to inexpensive, deep archive storage services in the public cloud.
- Send files from specified instruments for analysis on high performance compute clusters.
- Periodically delete files from the original instrumentation servers to free up space for incoming data.



**The institutes needed a way to offload manual tasks from researchers while ensuring files and data for clinical use cases were stored appropriately.**



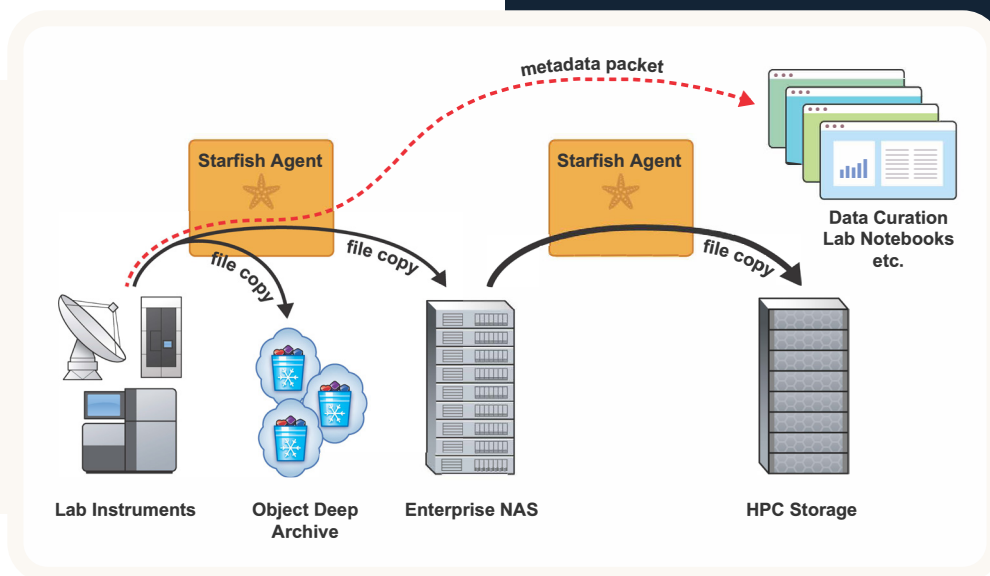
## The Solution

# STARFISH

Starfish is well-suited for bespoke file-based workflows. The Starfish scale-out jobs engine runs pre-built or custom scripts and records everything it does to a file, along with anything it discovers about a file as key-value metadata in the central database. The output of one job can serve as input parameters and/or triggers for the next job, enabling data movement and data processing workflows based on conditional logic. The logic is easily expressed in any programming or scripting language including Bash, Python, and Go. There is no need to learn another language or learn a complex system.

The research hospital engaged Starfish Professional Services to implement an automated workflow solution. Their tailor-made workflow performs the following tasks:

- Manages the hospital-wide instrument list, including parameters for storage locations and retention requirements.
- Transfers new files from instrument servers across storage mediums using MD5 hashes to validate file uniqueness and uphold data integrity.
- Directs files to specific locations based on the associated research group and the instrument type.
- Automatically removes older files from the instrument servers after a specified time period. This is only done after ensuring that valid duplicates are properly stored in central servers and backed up in offsite cloud services.

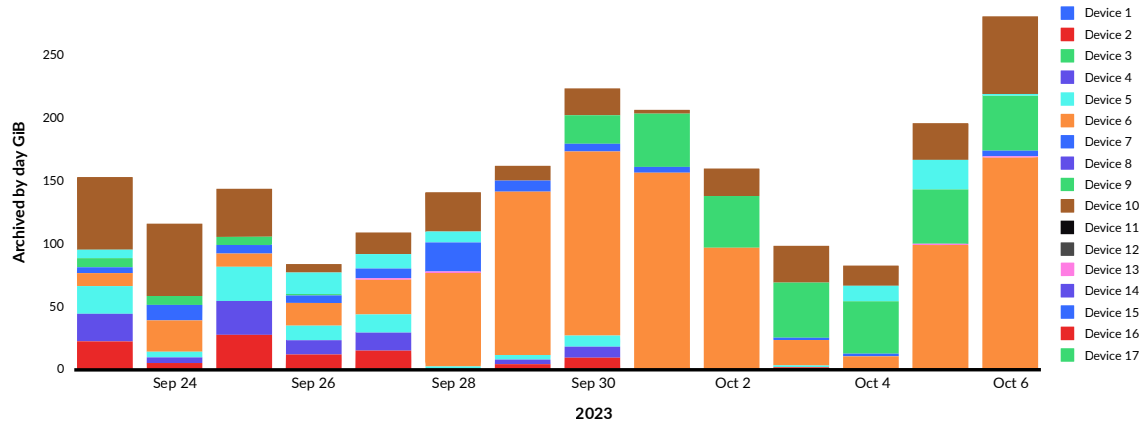


The client deployed the workflow to manage dozens of instruments across two labs. Each day, the custom workflow copies as much as 250 GiB from instruments onto storage devices throughout the environment.



## The Results

With the automated workflow in place, the research hospital ensures that instrument data is moved and stored according to strict clinical policies while enabling lab scientists to focus their energies on the science. The data is protected, and is where they need it, when they need it for their research.



Daily processed data across dozens of instruments



Once we implemented Starfish, complaints from researchers about managing data went to nearly zero... having the Starfish GUI front end to manage the different pieces and parts of the environment is essential.

— Sr. Programmer Analyst

## About Starfish Storage

Starfish is a unique software application for managing unstructured data at very large scale. Starfish combines a file system metadata catalog with a parallelized data mover and batch processor. You make discoveries and reports using the catalog. You move data fast and furiously and take other actions using the batch processor. The software is agnostic to storage vendors and works great with HPC file systems, enterprise NAS, object stores, and archives. Starfish allows your users to participate in storage management. Use cases include ROT cleanup, duplicate detection, backup, archiving, reporting, chargeback, content classification, and much more.