

# Chan Zuckerberg Biohub uses AWS DataSync to automate file transfers for in-cloud processing

## Challenge

Chan Zuckerberg Biohub runs genome sequencing jobs throughout the day. The resulting datasets need to be copied into Amazon S3 for processing and analysis, without disrupting on-premises workflows.

## Solution

Chan Zuckerberg Biohub uses AWS DataSync to quickly, securely, and cost-effectively copy sequencer data from their on-premises NAS storage to their S3 buckets. With the data in AWS, it can be processed efficiently using genomic analysis software running on Amazon EC2 instances.

## Benefits

With built-in scheduling, DataSync enables Chan Zuckerberg Biohub to easily run data transfer jobs on a daily basis. Using the filtering capabilities of DataSync, Chan Zuckerberg Biohub has configured their tasks to copy only the data generated each day, minimizing impact to on-premises storage systems and speeding data transfers.



Company: Chan Zuckerberg Biohub  
Industry: Life Sciences  
Country: United States  
Website: [www.czbiohub.org](http://www.czbiohub.org)

## About Chan Zuckerberg Biohub

Chan Zuckerberg Biohub actively nurtures and creates opportunities for leaders in science and technology to come together and drive discovery, setting the standard for collaborative science. Chan Zuckerberg Biohub conducts research that helps solve big health problems. By sparking collaborative, interdisciplinary work, these leaders empower the pursuit of intuition and the opportunity to explore the next questions — and answers.



We have genome sequencers running daily and need to transfer files a minimum of two times each day from on premises to Amazon S3 in an accelerated and efficient manner, in order for us to analyze and process the sequencer data in AWS. AWS DataSync serves this purpose perfectly as we are able to easily schedule and automate the file transfer process, avoiding the need for costly commercial transfer tools.



Lekha Karanam, Software Engineer - Chan Zuckerberg Biohub

