

# Labs of the Future: Enabling Integrated Research Labs with Cloud Technology



Scientists in leading research labs across the world struggle to store, analyze, integrate, and share data sets simply. Data is created from a wide variety of equipment types and manufacturers, and in many cases, that data is only stored on local servers. The majority of these labs are not set up to optimize the researchers' workday or streamline repetitive and time-consuming processes. Yet, with innovative thinking and the help of technology-based solutions, some research environments are addressing and solving these problems, creating a fully integrated 'lab of the future' that is data-driven and optimizes researchers time spent on scientific experimentation.

Establishing a digitized lab helps protect the valuable data researchers generate and can accelerate the research and discovery process by leveraging new technologies like machine learning. A lab of the future would enable automated data backup and archiving, fast data analysis, integration of artificial intelligence

to reduce the amount of time researchers spend on tedious tasks, and secure collaborations with other researchers internally and externally.

Industry figures show that over a 10-year period beginning in 2005, over 9,000 new partnerships were formed between pharmaceutical companies, with the goal of collaborating to innovate new drugs more quickly.<sup>1</sup> Seamless integration with external partners, while protecting organizational intellectual property (IP), will be vital to making these partnerships successful. Amazon Web Services (AWS) is making integrated research labs a reality through automated data storage from legacy equipment, secure data sharing to facilitate collaboration, and implementation of standardized data formatting. With data securely in the cloud, AWS makes it simple to collaborate with global research teams, utilize on-demand high-performance computing for analysis, and implement artificial intelligence or machine learning to speed time to actionable insights.



## THE CHALLENGE OF GETTING AND STORING DATA

Scientists in leading research labs struggle with everyday disruptions due to a lack of technology integration. Even in the most advanced labs, valuable experimental data may only be stored on PCs or servers attached to research equipment, leaving it vulnerable to being compromised or deleted. This lack of protection may be due to the organization failing to realize the value of its proprietary data. For research labs to be successful, data should be viewed as a new form of currency, which can continue to gain value as new data streams are developed, and can be used as a historical view to test new hypotheses against. Securing this data is an investment in the future value it can offer to your research, and your organization as a whole.



Housing data in the cloud can create automatic backups, taking the responsibility for archiving data away from the researcher.

Moving data to the cloud is an obvious solution, but few instruments are built to be cloud native, presenting a challenge for scientists who hope to easily assemble and integrate data. AWS DataSync offers a way around this roadblock, ensuring that important experimental data is automatically copied from the file systems that instruments write to, and securely store and archive it in the AWS Cloud. With the data in the cloud, researchers can begin to analyze their critical research in minutes, and can even automate workflows to initiate analytics or archival tasks as soon as the data arrives.

“Without the data in AWS, there’s no way we could innovate as fast. The AWS portfolio of hybrid storage and transfer services works with my existing lab computing environments and processes and helps us get our irreplaceable data safely into AWS, where our scientists can use it for whatever they need: machine learning, data analytics, or HPC,” said Lance Smith, Director of Research Computation at Celgene.<sup>2</sup>

Not only is data stored more securely in the cloud, but storing data in this way also prevents accidental loss due to equipment malfunction. Data stored on the instruments or local servers may not be backed up for an extended time, and power failure or damage to these systems can result in loss of valuable data. Housing data in the cloud can create automatic backups, taking the responsibility for archiving data away from the researcher.

Looking forward, new instruments on the market are working to integrate data seamlessly, including:

- BaseSpace Sequence Hub from Illumina, which allows an encrypted flow of data from the instrument to an AWS Cloud-based app to make collaboration and data processing faster and easier.<sup>3</sup>
- Thermo Fisher Connect, which connects lab equipment, automates data backup, integrates with lab management software, and can alert scientists working remotely.<sup>4</sup>

Another common challenge researcher’s face is to integrate data taken from instruments made by a range of manufacturers. Produced in a disparate

2 Duso, Wayne, “Expanding AWS Hybrid Cloud Capabilities with Block Storage on Snowball Edge.” *AWS Storage Blog*, April 29, 2019. <https://aws.amazon.com/blogs/storage/expanding-aws-hybrid-cloud-capabilities-with-block-storage-on-snowball-edge/>

3 Illumina BaseSpace Sequence Hub. Available at <https://www.illumina.com/products/by-type/informatics-products/basespace-sequence-hub.html>

4 Digital Science | Thermo Fisher Scientific US. Available at <https://www.thermofisher.com/us/en/home/digital-science.html>

30%



The potential for innovation through partnership is already well known in the industry, with figures showing that 30% of the drugs currently in development at large pharmaceutical companies were initially developed by another company.

range of formats, the data that comes from each tool can be hard to integrate and analyze without preparation work to create a uniform format across each data set.

“Most desktop analyzers are manufactured by different vendors, and they all produce test result data in different formats,” said Savitra Sharma, CEO of a research analytics firm. “In order to combine that data, the lab tech must manually massage that data and put it into one uniform format. This data then needs to be manually uploaded to a laboratory information management system (LIMS).”

AWS and APN Partners, like TetraScience, facilitate data integration from diverse data sources and standardize data into vendor-agnostic and analytics-ready formats, eliminating the data silos found in traditional research labs.<sup>5</sup> This helps researchers more effectively focus on data analysis, rather than reformatting data that has been aggregated from a variety of instrumentation.

### GLOBAL DATA SHARING AND ADVANCED ANALYSIS

Collaboration through global data sharing is key to fueling discovery, and both pharma and public research entities are embracing this idea by investing more in collaborative efforts. The potential for innovation

through partnership is already well known in the industry, with figures showing that 30% of the drugs currently in development at large pharmaceutical companies were initially developed by another company.<sup>6</sup>

However, as the value of data as a form of currency becomes more evident, sharing sensitive research data requires extra caution to ensure that it remains secure and IP is protected. Data-sharing methods using flash or hard drives are not scalable to the large data sets now common in research, are impractical for long-distance collaborations, and also introduce data security risks. Cloud-based data storage provides a cost-effective, scalable, and secure modern alternative. And once data is securely moved to the cloud, AWS can facilitate multinational collaborations by enabling the primary organization to create isolated access for researchers from different organizations. AWS is being used to power many large research consortiums, such as CHARGE,<sup>7</sup> in which more than 200 scientists from five global institutions work together to try to identify genes that may be involved in the development of heart disease. With more than 450 TB of data in play in the CHARGE project, “having to ship out hard drives to so many people would be a logistical nightmare,” Narayanan Veeraraghavan, lead programmer scientist at Baylor, says. “Data would have to be encrypted

<sup>5</sup> Amazon Web Services. “Data Integration from Cloud-Enabling Lab Instruments.” YouTube, YouTube, Feb. 6, 2019, [www.youtube.com/watch?v=F6lStXFBjl](http://www.youtube.com/watch?v=F6lStXFBjl).

<sup>6</sup> Buvailo, Andrii. “Pharma R&D Outsourcing Is On The Rise.” *BioPharmaTrend*, Aug. 13, 2018. [www.biopharmatrend.com/post/30-pharma-rd-outsourcing-is-on-the-rise/](http://www.biopharmatrend.com/post/30-pharma-rd-outsourcing-is-on-the-rise/).

<sup>7</sup> Baylor Case Study—Amazon Web Services. Available at <https://aws.amazon.com/solutions/case-studies/baylor/>





at all points. With so many scientists handling so many hard drives, there would be a lot of failures, because not everyone would be able to follow the security guidelines.”<sup>7</sup>

Cloud-based collaborations also have the additional advantage of being cost-effective for research collaborators with limited resources. The prohibitive cost of investing in better computers and the infrastructure to support them can limit valuable contributions from collaborators and contractors who need to move quickly or have limited funding. The on-demand computing power of the cloud can be at the fingertips of every contributor, regardless of location.

Powerful cloud computing is a key component that enables collaborating researchers to run hundreds or even thousands of programs, calculations, or simulations in parallel and get fast results. Companies like Celgene, Bristol-Myers Squibb<sup>8</sup>, and Fabric Genomics<sup>9</sup> are already using the AWS cloud to make

high-performance computing possible to help them speed research or clinical analysis.

### INTELLIGENCE THAT CAN DRIVE ACTION

As integrated data pipelines become standard in advanced research labs across the world, moving from raw data to insights and action can happen much more quickly. With data more securely accessible in the cloud, additional technology, such as machine learning, can be incorporated to complement the work of scientists and researchers. Technology that can analyze data, predict trends, alert researchers to issues that warrant further analysis, or perform simple but necessary daily tasks makes lab environments more efficient.

Programs like LabAlert, developed by Celgene,<sup>10</sup> continuously monitor instruments and application status so scientists can focus on other value-added tasks. Using Amazon Simple Notification Service (SNS) and AWS Lambda, researchers are notified on their mobile devices if there is an equipment malfunction or if an experiment finishes while they are away,

<sup>8</sup> “AWS Case Study: Bristol-Myers Squibb.” Amazon, Amazon, <https://aws.amazon.com/solutions/case-studies/bristol-myers-squibb/>

<sup>9</sup> “Fabric Genomics Case Study - Amazon Web Services (AWS).” Amazon, Amazon, <https://aws.amazon.com/solutions/case-studies/fabricgenomics/>

<sup>10</sup> Bookbinder, Maxine. “Celgene’s LabAlert System Focuses On Scientific Workflow Efficiency, User Experience.” *Bio-IT World*, June 18, 2018. <http://www.bio-itworld.com/2018/06/18/celgenes-labalert-system-focuses-on-scientific-workflow-efficiency-user-experience.aspx>

thereby making data collection more reliable and less reliant on manual intervention.

Similarly, AstraZeneca<sup>11</sup> uses machine learning on AWS to help automate some of the most tedious portions of tissue data labeling, which helped them reduce the time their researchers spent cataloguing samples by 50%, thereby allowing scientists to focus more of their time on valuable research activities.

### WHAT DOES THE FUTURE HOLD?

Cloud-based technology can provide streamlined ways to access, standardize, and share instrumentation data, while offering cost-effective and powerful computation and analytics capabilities. And as artificial intelligence and machine learning are further incorporated into modern research labs, additional monotonous tasks can be offloaded from scientists to let them focus on

---

11 *Wired Insider*. "Machine Learning Is Transforming Drug Discovery at AstraZeneca." *Wired*, July 2, 2019, [www.wired.com/brandlab/2019/07/machine-learning-is-transforming-drug-development/](http://www.wired.com/brandlab/2019/07/machine-learning-is-transforming-drug-development/).

value-added research activities. Future innovations may include:

- Use of machine learning to accelerate cataloguing, analyze or prioritize experimental results, and suggest a future course of action.
- Expanded use of robotics and automation to allow for batch-of-one precision medicine manufacturing.
- Digitization and data integration across entire organizations to connect typically siloed departments, from research to manufacturing and commercialization.

Seamlessly integrating research, facilitating collaborations, and incorporating machine learning as part of a 'lab of the future' will enable pharmaceutical researchers to advance the pace of science and medicine daily. Changing the way data is collected, stored, and shared is the clear way to move forward in an industry that demands global collaboration to power the research that can one day address important human health issues.

---

For over 13 years, Amazon Web Services has been the world's most comprehensive and broadly adopted cloud platform. AWS offers a wide set of services including compute, storage, databases, analytics, networking, machine learning, management tools, IoT, security and enterprise applications, from 69 Availability Zones within 22 geographic regions. AWS is trusted by millions of active customers around the world—including the fastest-growing Life Science startups, largest enterprises, and leading government agencies—to power their infrastructure, make them more agile, and lower costs.

**To learn more about AWS, visit <https://aws.amazon.com/health/biotech-pharma/>**

---