Barriers to Academic Use of Commercial Cloud Labs

C Armer[1], S Golas[2], T Kalil[3] and E DeBenedictis[4]*

[1]Department of Automation, Univeristy of Washington, USA

[2]Department of Automation, Massachusetts Institute of Technology, USA

[3]Department of Automation, Schmidt Futures, USA

[4]Department of Automation, Bioautomation Challenge, USA

***Corresponding author:** Erika Alden DeBenedictis, Department of Automation, Bioautomation Challenge, USA.

Summary

Cloud laboratories have the potential to improve scientific reproducibility, allow experiments to be easily scaled once developed, and enable robust closed-loop engineering workflows in experimental life science. However, although industry has used cloud laboratories for nearly a decade, the same technologies are rarely used in academic science. Here, we explore the barriers to entry that prevent basic science from benefiting from 'programmable experiments.' We highlight possible avenues to kickstart broader use of these technologies.

Introduction

Widespread cloud lab usage would allow experiments to be shared the same way that open source software is shared today. This allows experiments to be specified in a standardized format, with all details included, in script form. Experiments can be easily version controlled, repeated, transferred between labs, or scaled up once developed.

Cloud labs also offer access to a wider array of equipment than any lab will ever have locally. They eliminate the up-front cost associated with purchasing equipment, the cost of space, and the cost of long-term service contracts. These facilities give academic groups instant access to a wide variety of specialized instruments that may not be available at their local institution, effectively democratizing access to these instruments and enabling less-well resourced groups access to top-of-the-line equipment.

Despite these advantages, academic users today rarely take advantage of cloud labs. We explore two major barriers to academic use. First, pricing is designed for industry customers, and is not compatible with the normal size of academic budgets or available funding opportunities. Second, resources for effectively using cloud labs, including software and training materials, are siloed and inaccessible to academics. We assess possibilities by focusing with the two major cloud lab providers in the US, Emerald Cloud Labs and Strateos, as examples. We explore strategies to remove both of these barriers to enable wider use.

Barrier #1: Pricing Models are Designed for Industry Customers

The two primary cloud lab providers in the US, Emerald Cloud Lab and Strateos (formerly Transcriptic), both cater primarily to customers in industry. Their pricing models are designed with industry as the primary customer. Providers often aim for their pricing to compare favorably for startups who would otherwise need to pay separately for bench space. In contrast, most academic groups who are interested in cloud labs already have access to bench space and pay overhead for this access, effectively double-

charging them for facilities access if they also use a cloud lab. Together, current pricing models are not attractive to academic customers.

Subscription service pricing model

Emerald Cloud Lab is the provider that is closest to being suitable for academic use because they encourage users to develop new methods themselves. Emerald Cloud Lab has a 'subscription service' model, in which the user pays for access to the system, and then may pay additional fees for reagents or robot time. Emerald offers two pricing models:

• **All-inclusive pricing**, in which the user pays a large subscription fee for access to a substantial amount of robot time each month, without additional per-hour equipment fees. This pricing level costs $25k/month, with a 1 year minimum commitment, for a total entry cost of $300k.

• **Pay-as-you-go pricing**, in which you pay a relatively modest subscription fee (~$2k/month), and then separately pay-per-hour for time on each instrument. This pricing model is not attractive for academic labs that are first experimenting with remote experiments because developing cloud lab methods is often resource intensive, making it possible to spend even more than $25k in a month. Additionally, the per-experiment mindset is not conducive to method development, which inherently requires a lot of trial and error.

We will focus on the all-inclusive pricing moving forward.

Specific method automation

Strateos takes a different approach and offers to automate specific methods and conduct them at scale. In this approach, the user does not perform the automation engineering (or any programming) themselves. Strateos charges around $10-$15k to automate a method, and then requires a contract to conduct that assay at scale on the order of $10k/month for a year minimum. This brings the total cost to entry to $130k, and restricts the user to a single automated method.

Barrier to entry is cost and risk

In both cases, the cost to entry is high ($300k for general access, or $130k for a single method), and the contract lengths are long (a year minimum). This is a very large expense, and a large risk, especially when the cloud lab model is largely unknown and untested, and an academic may not be sure whether or not a project they have in mind is well suited to the cloud lab model.

Even assuming that an academic lab chooses to move forward, there are currently few funding sources designed to support cloud lab access for academics. Very few labs have enough discretionary funds to cover this cost. Equipment grants are unlikely to be willing to pay for cloud lab access. Although some have succeeded in putting

cloud lab access in budgets on an R01 or equivalent, it is generally not obvious whether or not the NIH or other funders would look favorably on this sort of expense in the budget, considering they have no stated policy about cloud labs. Additionally, it is difficult to seek out funding without already having data demonstrating the suitability and utility for a particular application. For these reasons, it is risky to spend time searching for a funding source to pay for this non-traditional style of science.

Suggested solutions

Academic users will be able to experiment with cloud labs if the cost and risk to entry is decreased. One model is for universities to provide cloud lab access. In the academic system, access to facilities is usually handled at the University level, so this model is likely the natural long-term solution to cloud lab access as this technology becomes more widespread. Carnegie Mellon University has recently experimented with this model. Here, the university department sponsors a cloud lab account, negotiates a bulk rate with a cloud lab provider, and offers access, training, and community support to its members. CMU is now building a local facility to further reduce costs and customize the equipment they offer.

Shorter term, Universities will require experience with cloud lab providers before they prioritize creating access programs. A second model for decreasing cost and risk is the 2022 Bioautomation Challenge, a grant program that provides cloud lab access. In this program, labs propose a project that would benefit from automation. Selected proposals receive short-term access (3 months) to a cloud lab account as a 'trial period,' followed by an optional 9 months of access to build out their project. By the end of the year, the academic lab has experience with the product, an accurate assessment of how much it costs to conduct their particular experiments robotically at scale, and initial data with which they can apply for continued grant funding. Entry into a long-term contract could occur after this initial proof-of concept period. Programs like this are a strategy to rapidly stress-test cloud labs and expose this technology to academic groups at many universities.

Cloud lab providers are generally interested in participating in these programs. As academic users become a larger fraction of their customers, they may create pricing models that are designed for academic customers that may include these same essential features, notably including a trial period and reduced cost during initial data acquisition.

Barrier #2: Resources are Siloed

Automation engineering of experimental biology requires a unique set of multidisciplinary skills, making it challenging for students to succeed without intentional training. Traditional teaching siloes knowledge of the several disciplines that are required, including experimental biology, chemistry, computer science, systems engineering, and robots. We must create new

training and teaching resources to enable students to use cloud labs to their fullest extent.

Existing resources

Of the two cloud lab providers in the US, Strateos sidesteps this issue by having all automation engineering be performed by professionals. This model makes Strateos more of a 'robotic CRO' and less of a company that provides access to 'programmable experiments' broadly. Emerald Cloud Lab intends for the customer to perform their own automation engineering, making it more suitable for training students who will be the next generation of automation engineerings. This company provides new users with a brief industry-oriented training and certificate program, which is not sufficiently in depth for most students.

Although in principle based on code, today very little open source software is available for cloud labs. The lack of software means that every new user needs to start from scratch.

Notably lacking are open source libraries that allow users to rapidly customize standard workflows for their own use cases. entire workflows are not available as open source libraries, meaning everyone needs to start from scratch.

Suggested new resources

New strategies for pedagogy of automation engineering are required. CMU is developing a new interactive course that will train students <>

Creating an online community is essential for rapid software development. Today, industry customers who have expertise in cloud lab programming do not interact with one another or with academic customers. In the future, open source software and public question-and-answer forms like stackoverflow will enable these groups to interact and share code and expertise.

Conclusion

This program is an opportunity to improve the reproducibility of science in academica broadly, and may be sufficient to encourage University departments and other funders to offer additional funding opportunities. It is also an opportunity to gather information from participating labs about what the onboarding process is like, the experience of transitioning from doing experiments in person to remote automated experiments, and other cultural and psychological aspects of doing version-controlled biology experiments that are relevant for understanding the prospects and challenges of achieving reproducible life science broadly.

Acknowledgement

None.

Conflict of Interest

No Conflict of interest.