

Why Use A Cloud Laboratory



**Advantages of a Cloud Laboratory
Over the Traditional Laboratory**

Traditional Labs Present Barriers to Experimentation

Life scientists in all sectors face substantial and persistent logistical and labor barriers to running experiments

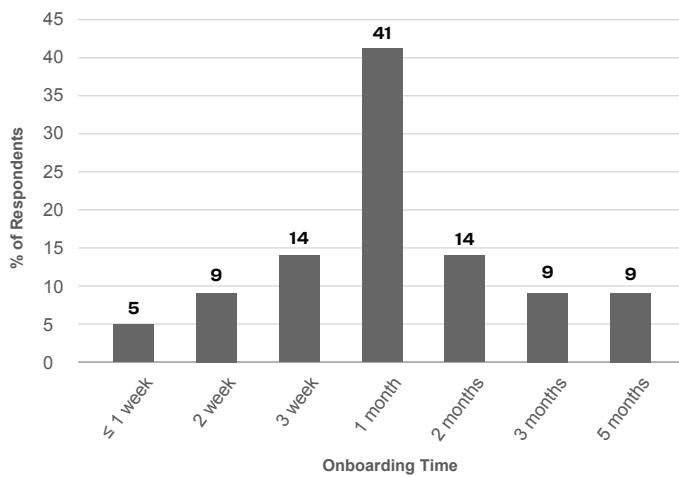
Equipment Access Problems



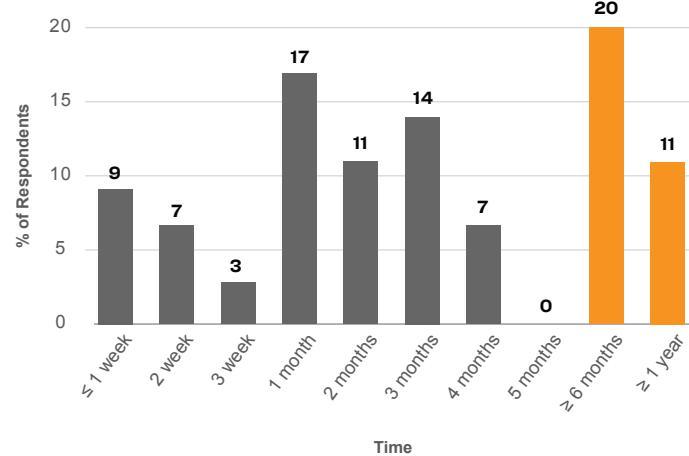
- **Over 80%** report access to instrumentation as a major impediment
 - Many labs lack full instrumentation required in their research
 - Maintenance and downtime of available instruments are additional barriers
 - Most popular work around (73% of respondents) was to request personal favors from scientists in other labs
- Average onboarding time for each new experiment type is about a month, creating barriers to conducting new or unfamiliar experiments
- Lengthy troubleshooting can stymie onboarding new experiments for months
 - Unpredictable, with a broad distribution ranging from weeks to years
 - **Over 30%** of respondents said they had suffered a delay greater than six months on a project

Emerald customer survey, n=130 Scientists from across the industry (mix of startups, large companies, and academia) (2013)

Average Experiment Onboarding Time



Troubleshooting Time



With a Cloud Lab, Every Instrument is at Your Fingertips

- **Science, not instrument availability, drives decisions**
 - Full Control of over 150 unique scientific instruments
 - Complete list in Scientific Instrumentation section
- **Minimize time from idea to experiment**
 - Experiments start within 24 hours from the moment you submit them
 - No onboarding time for the 150+ instruments already online and waiting
- **Drive all instruments remotely through a software interface**
 - Instrument and equipment shared at maximum efficiency, with jobs automatically distributed throughout the facility
- **Never fight with broken instrumentation again**
 - All instruments operated at GxP-level standards with scheduled maintenance and qualifications
 - Dedicated team of scientists on-site to manage any instrument issues that arise
- **New instrumentation developed by the CMU Community can be deployed in the cloud lab for wider shared use**
 - Researchers can set up their novel equipment in the CMU cloud lab facility, and then use Emerald's developer tools to build a computational interface to this equipment allowing any other researchers on the cloud to conduct experiments using that equipment on demand, 24/7/365
 - Maintenance and qualification of the instrumentation can then be programmed as well to automatically keep the equipment in top working order at all times



In Traditional Labs, Science Can Take a Backseat to Labor

- More time is spent on manual labor than on all the intellectual aspects of science combined (designing experiments, developing theories, building models, analyzing data)
- Creative ideas require immense labor to execute. This favors the most well-funded teams rather than those with the most ambition
- The physical effort involved in lab work creates a barrier to scientific exploration because it is incredibly labor-intensive to pursue new research initiatives
- It is hard to incorporate new workflows as training or capital investment is often too large of a barrier



With a Cloud Lab, Science is in the Driver's Seat

- **Focus on the science**
 - Scientists using a cloud lab spend their time designing experiments, analyzing and interpreting experimental results, and focusing on new ideas
 - Experiments in a cloud lab are run automatically on-demand and are controlled entirely using software — no physical labor involved
 - Scientific talent can conceive of, execute remotely, and analyze data from more experiments (3x to 5x) than in a conventional setting.
- **Increased throughput through automation**
 - Scientists working on the ECL daily have historically processed over 50,000 samples per year per person
 - Experiments run nonstop, even while you sleep and on the weekends; facility operates 24/7/365
 - Organizations using cloud labs do more with the same, with documented productivity improvements in the 300% to 700% range
- **Greater interdisciplinary collaboration**
 - Break down barriers between disciplines by making all the instrumentation available to everyone on the system
 - Methods transfer seamlessly between groups, protocols are repeated at the push of a button
- **Inventory and resources managed automatically**
 - The CMU cloud lab keeps frequently used materials in stock
 - Easily place and track external orders anytime

In Traditional Labs, Experimental Reproducibility Presents Enormous Challenges

The image below highlights an example of the types of “missing” information from just a small portion of a method section in a paper published in the Journal of the American Chemical Society

Criteria for sequence selection?

Container type? With what rotor?

Stirring? Storage conditions?

Selected how? Agitation?

Prepared how? Resin type?

Single nucleotide polymorphism analysis. An 18 μ L slurry of ~7.5 μ M solid-phase translator (HCV, Output^{*}), 3 μ L of 30 μ M HCV[†] and 20 μ L of IDE buffer were allowed to incubate at room temperature for various incubation times. After incubation, the sample was centrifuged at ~1,000 RPM briefly and 25 μ L aliquots of the supernatant were carefully removed. Those aliquots were then injected into the cuvette and emission recordings were taken with one second integration time for >1,000 seconds. Ten second running averages were then applied to smooth the curve.

Wavelength? Left, center, or right averaged?

Removed how? Cuvette material?

Aspiration rate? On what instrument?

Syringe material?

- Results are most easily reproduced when conducted by the same scientist working in the same laboratory. Reproducibility of experiments don't transfer well to other scientists and laboratories. Why is this the case?
- Ambiguity in communication of methods and data analysis is at the core of the reproducibility problem
- With traditional method descriptions it's impossible to know *a priori* if the information presented is sufficient to reproduce an experiment without actually running it
- Routinely reproducing experiments is time-consuming and resource intensive and slows the overall pace of progress

Inconsistent Results

71.1%

Nature Reviews Drug Discovery 10, 712 (2011)

Inconsistent Results

88.5%

Nature 483, 531-533 (2012)

Experiment > Experiment 12

Model[Instrument, HPLC, "UltiMate 3000"]

Preparative

IonExchange

{Temperature, Pressure, UVVis}

Model[Item, Column, "DNAPac PA200 4x250mm Column"]

45 Celsius

Forward

None

Forward

Model[Sample]

Model[Sample]

Model[Sample]

Wavelength

Temperature

Volume

FlashSolution

Time

TimeA

TimeB

TimeC

TimeD

FlowRate

Gradient

Detector Parameters

AbsorbanceWavelength

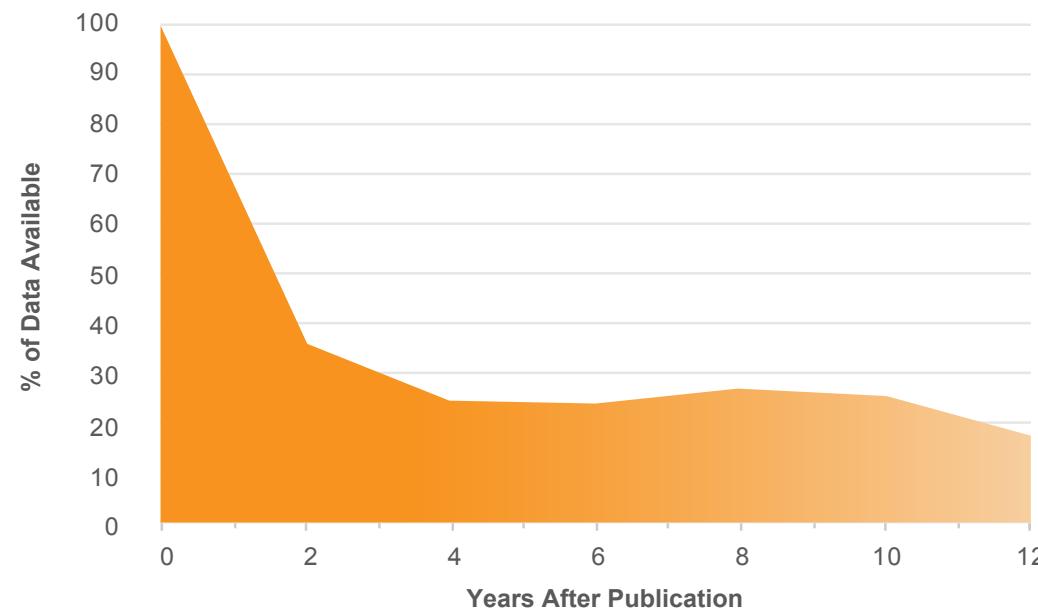
FractionCollection

With a Cloud Lab, Experiments are Codified

- All experiments in the cloud lab are driven by an enumerated instruction set, much like a microprocessor
 - Just as a microprocessor can reproducibly execute a defined series of steps, the cloud lab can reproducibly conduct any experiment
- Push-button reproducibility
 - Any protocol run historically in the cloud lab can be re-run identically with fresh samples on-demand
- 360° data capture — no data left behind
 - The cloud lab automatically collects all data and protocol steps that are scientifically and operationally relevant
 - In addition to primary data, ancillary data such as full sample trace history, instrumentation information (including calibrations and qualifications), and metadata captured through embedded sensors in the facility are recorded
 - The linked data network ties together all aspects of an experiment and analysis together in a scale-free network
- Easily transfer methods between people or groups
 - Sharing experiment IDs makes it easy to collaborate with scientists across the room or across the world
- Codifying technique turns all errors into systematic errors
 - Easier to identify and make systematic fixes that ensure issues do not reoccur

In Traditional Labs, Availability of Primary Data Drops Off Precipitously Over Time

Less than 40% of Literature Data Accessible After Just Two Years



Current Biology 24, 94-97 (2014)

Because much of the relevant information is either uncaptured or unstructured, personal communication with the original authors via email or phone calls is often necessary to make sense of any raw source files (if they can even be located).

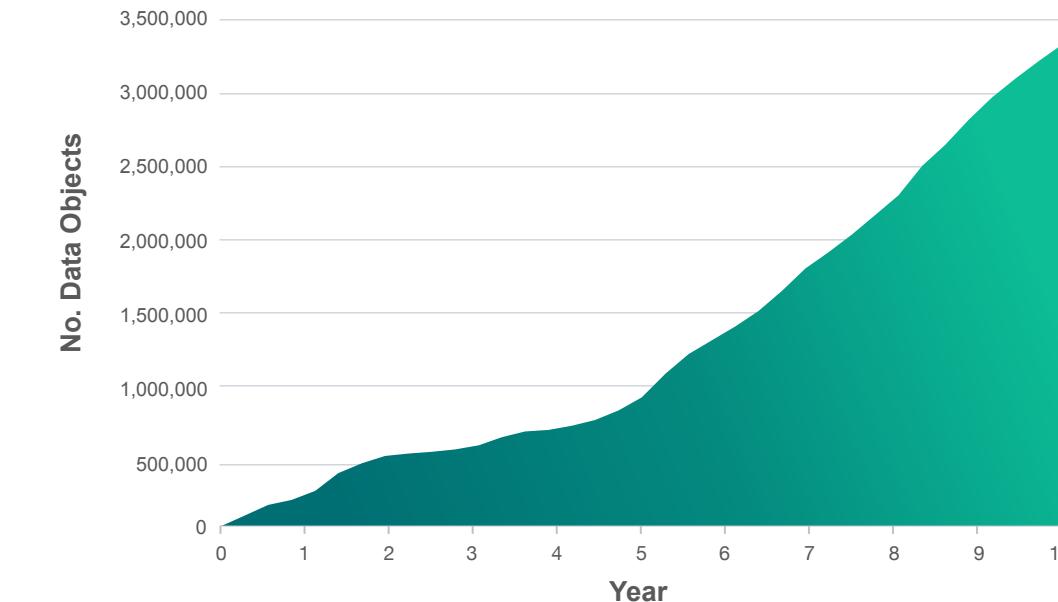
- Tremendous human and societal effort goes into generating data that tragically fades rapidly over time. Data is lost in paper notebooks, hard-to-access files, and rigid data systems.
- Difficult to build on previous work, creating continuity challenges across generations of a lab
- Combining new and old data sets is difficult when primary experimental data is not available; even when available, method data is often missing or incomplete
- Machine learning is stymied by missing information or data that's not reproducible and not well-structured
- Lost opportunity for meta-analyses and other higher level conclusions

With a Cloud Lab, Data Generated Compounds in Value Over Time

A laboratory, company, or institution conducting day-to-day experimentation in this highly structured and connected system would produce a dataset with detail, sophistication, and complexity without historical precedent.

Imagine what you could do with just a year of all of your experimental data indexed and searchable on the web!

Scientific Data Generated on the ECL



Data accumulated using the cloud lab always remains just milliseconds away from download on the cloud. Moreover that data is already highly structured, indexed, searchable, and, most importantly, it is connected to the experimental techniques that generated it (with push button reproducibility).

- All data captured digitally and automatically
 - Eliminates need for paper lab notebooks and printouts
 - No more efforts wasted in data transfer to ELN
- Everything accessible on the cloud to all users with valid credentials
- Methods valid and reproducible for years after initial execution
- Data is automatically structured, indexed, and made quickly searchable for instantaneous retrieval
- Standardized data ontology amenable for data mining and machine learning
- All data is traceable and linked to its source techniques and lab notebook context
- Data gathered on enterprise accounts compliant with FDA data retention and access policies