Database, Methodological Tools, and Research Opportunities: Creative Destruction Lab and Early-Stage Technology Ventures*

Extended Abstract
Please Do Not Distribute

Amir Sariri† Evgenia Gatov‡ Kyle Robinson§ Avi Goldfarb¶

September 26, 2024

*We thank Ajay Agrawal, Kevin Bryan, Joshua Gans. Financial support from the Government of Canada's Strategic Innovation Fund is acknowledged. Creative Destruction Lab supports the creation of the database described herein.

†Purdue University Daniels School of Business
‡Creative Destruction Lab
§Creative Destruction Lab
¶Rotman School of Management & NBER

1 Introduction

The early stages of firm formation remain one of the least understood aspects of entrepreneurship. Decisions made during these formative periods often have irreversible consequences, shaping the trajectory and potential success of new ventures. However, understanding these decisions has been hindered by a lack of data that captures the real-time evolution of startups. Startup programs offer a unique window into this process, providing detailed records of the early development of technology ventures, including mentoring interactions, business objectives, and company progress.

In this extended abstract, we introduce a new database built from one such program called the Creative Destruction Lab (CDL), a global science-based entrepreneurship with over $30 billion of equity value created by its participating firms. CDL offers researchers the opportunity to study the early decision-making processes of startups as well as the foundations of advice.

Since its inception in 2012, CDL has grown substantially, encompassing multiple streams across various technology domains and admitting thousands of startups from around the globe. The program's structured mentoring sessions and competitive selection process provide a rich environment to observe how startups prioritize objectives, experiment with market opportunities, and seek guidance from experienced mentors, such as angel investors and venture capitalists. By systematically collecting data on these interactions, the CDL data provides new opportunities to explore the early dynamics of firm development, the role of mentorship, and the impact of different approaches to business strategy.

In developing and providing this database, our objective is to create a resource that enables deeper insights into the innovation process by advancing our knowledge about how startups navigate the crucial early stages of development. Therefore, we believe that the scholarship opportunities created by this database will offer a foundation for research on technology entrepreneurship, the university-industry interface, and innovation processes.

The rest of this short report provides an overview of the CDL program, a high-level description of the available data, and information on how to access the data for research.

2 Creative Destruction Lab: An Overview

CDL is a nonprofit that operates in business schools and is steered by senior faculty. The essence of CDL is four in-person "sessions" every eight weeks to help founders prioritize three measurable business objectives to focus on at the expense of everything else. A fifth and final graduation meeting concludes the program year. Since its inception in 2012, CDL has grown from one stream at one school with 24 alumni, to 13 business schools across seven countries, with 23 specialized technology streams, and more than 1,000 alumni estimated to be worth over $28 billion.

Admission to CDL is competitive and open to startups from anywhere around the world. Candidates are subjected to a rigorous evaluation process that includes submitting a detailed application and participating in business and technical assessment interviews. Finalists are offered admission to a technology "stream" at a unique "site" (hereafter, stream-site pairs are referred to as "track"). The matching of startups to tracks is centrally administered via the two-sided Gale-Shapley deferred acceptance algorithm. Each stream assembles mentors with relevant domain expertise, such as prior investment history in the same technology domain. Technology streams include artificial intelligence, space, and quantum computing.

Business mentors are primarily angel investors and venture capitalists from established ecosystems such as Silicon Valley, Boston, and Toronto. As part of their agreement to participate in CDL, mentors cannot delegate their role to an associate or employee. Based on data from the 2018-19 cohort, each track has an average of 18 startups (SD = 4.8) and 19 mentors (SD = 6.8), with 75% of mentors participating in a single track, 18% in two tracks, and the remaining 7% in three or more tracks.

2.1 Mentoring Process

A week before each session, mentors in each track electronically receive updated one-page dossiers like the one shown in Figure 1 on every startup in their track. These dossiers outline the founders' proposed objectives for the upcoming eight weeks, the status of the previously finalized set of objectives, and updated financial details. The document also contains a hyperlink to a longer curated document with further details ranging from target customer and core technology to founders' educational back-

ground. The email asks mentors to familiarize themselves with each firm's progress and formulate their feedback on the proposed objectives.

On the morning of session days, founders meet privately with 4-6 mentors from their track to receive one-on-one feedback on their proposed objectives. In the afternoon, each track's mentors and founders convene in a classroom like the one shown in Figure 2 to debate and reconcile individual mentor feedback into a final set of objectives. A business school professor moderates these debates one startup at a time until each firm finalizes its objectives. Sessions conclude in the early evening with deliberations, whereby the debate moderator asks mentors to raise their hands if they feel equipped to support each venture in achieving its finalized objectives.

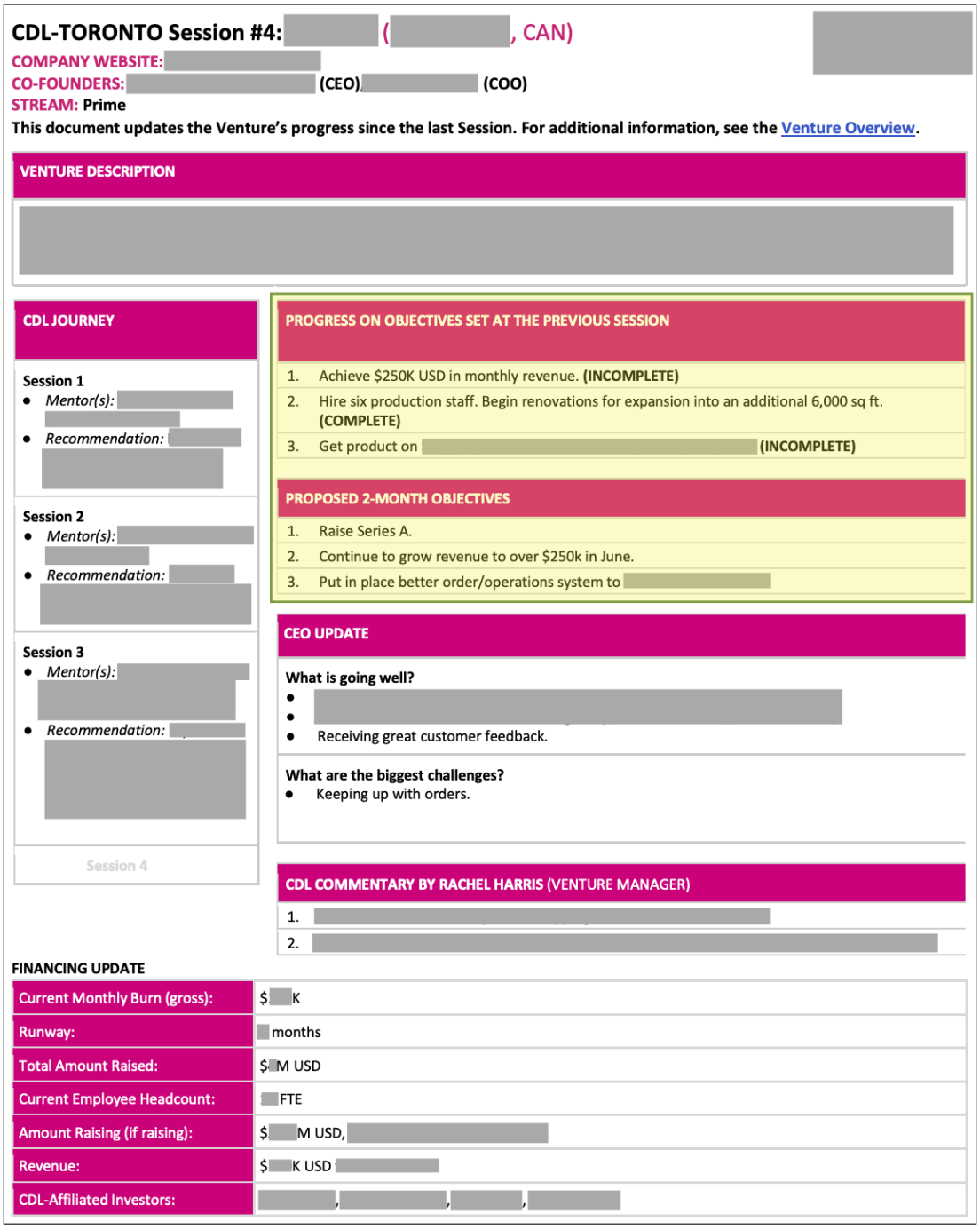
Mentoring decisions are costly as each obligates a mentor to commit four hours of their personal time to helping the startup achieve its objectives. The modal (average) startup receives one (1.61) mentor, and the modal (average) mentor selects one (1.64) startup. Decisions are also high-stakes for startups as those without formal support are dropped from subsequent sessions. SEP managers responsible for each startup connect the founders with their mentors shortly after the session and facilitate setting up the meetings. They also touch base with founders throughout the eight-week cycle to document progress on objectives and track mentors' honoring of their time commitment. The cycle ends with founders submitting a draft dossier for the next session.

3 A Brief Overview of Data

Data collected throughout routine CDL operations includes information on ventures, individuals, financing, process, direct communications between stakeholders, administrative records, and online resources.

At the venture level, data contains venture overviews, core technology, geography and industry. For companies admitted to the CDL program, additional information includes financial and non-financial updates regarding the progress of each company while they stay in the program. For individuals, CDL data contain information on educational and professional histories, which applies to both founders and mentors.

Figure 1: Example of Startup Dossier



CDL-TORONTO Session #4: [redacted] ([redacted], CAN)

COMPANY WEBSITE: [redacted]

CO-FOUNDERS: [redacted] (CEO), [redacted] (COO)

STREAM: Prime

This document updates the Venture's progress since the last Session. For additional information, see the Venture Overview.

VENTURE DESCRIPTION

[redacted]

CDL JOURNEY

Session 1
- *Mentor(s):* [redacted]
- *Recommendation:* [redacted]

Session 2
- *Mentor(s):* [redacted]
- *Recommendation:* [redacted]

Session 3
- *Mentor(s):* [redacted]
- *Recommendation:* [redacted]

Session 4

PROGRESS ON OBJECTIVES SET AT THE PREVIOUS SESSION

1. Achieve $250K USD in monthly revenue. **(INCOMPLETE)**
2. Hire six production staff. Begin renovations for expansion into an additional 6,000 sq ft. **(COMPLETE)**
3. Get product on [redacted] **(INCOMPLETE)**

PROPOSED 2-MONTH OBJECTIVES

1. Raise Series A.
2. Continue to grow revenue to over $250k in June.
3. Put in place better order/operations system to [redacted]

CEO UPDATE

What is going well?
- [redacted]
- Receiving great customer feedback.

What are the biggest challenges?
- Keeping up with orders.

CDL COMMENTARY BY RACHEL HARRIS (VENTURE MANAGER)

1. [redacted]
2. [redacted]

FINANCING UPDATE

Current Monthly Burn (gross):	$[redacted]K
Runway:	[redacted] months
Total Amount Raised:	$[redacted]M USD
Current Employee Headcount:	[redacted] FTE
Amount Raising (if raising):	$[redacted]M USD, [redacted]
Revenue:	$[redacted]K USD [redacted]
CDL-Affiliated Investors:	[redacted], [redacted], [redacted], [redacted]

Notes: This figure shows a sample startup dossier distributed to mentors before sessions. It includes updated objectives, a status update from the CEO, commentary by the CDL manager responsible for the startup, the latest financial information, and a history of main mentor recommendations from prior sessions. Portions that may reveal the identity of the startup or CDL are redacted.

Figure 2: Finalizing Objectives via a Moderated Debate



Notes: This image shows an in-progress discussion in the large room. Founders and mentors engage in a debate moderated by a business school professor (hidden behind the founder) to arrive at a finalized set of objectives for the next eight weeks.

Regarding financing, detailed financing records exist while companies are in the program, but CDL also tracks and records, if disclosed, additional funding rounds of the ventures after they leave the CDL program. These data include amounts raised, financing stage and instruments, as well as exact or estimated valuations.

CDL data also contains detailed records of the mentoring process and the support companies receive during the program. These data include business objectives of the companies and the progress they make on each, the time commitments mentors make to help them achieve those objectives, as well as meeting notes from debates and discussions that take place between mentors and founders on business objectives to prioritize and potential pivots to make.

To gather these data, we gather and organize information from several sources. These include the detailed application that ventures submit to be considered for admission into the program, but also direct communications with stakeholders to ascertain financing information such as financing events and valuations. In addition, administrative record-keeping such as scheduling data allows identifying the attendance of each individual at mentoring sessions. We also complement our internally-obtained data with online records such as company websites, commercial databases, and online personal profiles and resumes.

3.1 Sample Mentor & Startup Data

Since its inception in 2012 with a cohort of 18 startups admitted, the program has been consistently growing by expanding to new sites and adding new streams. By the 2021 cohort, there were 600 startups admitted from a pool of nearly 6,000 applications.

As a glimpse into the diversity of variables in CDL data, here we draw on a recent project by Sariri (2024) who uses data from the 2018-2019 cohort to characterize the key elements of advice and document systematic differences between angel investors and venture capitalists in the provision of mentoring support. In terms of mentors, after attrition, there are 192 business mentors comprising of 44 angel investors and 148 venture capitalists. Table 1 summarizes these mentors characteristics. Angels have twice the operating experience of VCs, with all being former founders and 61% having had an exit. In terms of other types of experience, however, angels and VCs are similar: both have significant managerial experience, and do not differ significantly in terms of technical and academic jobs held. Educational background is also balanced across majors and highest degrees earned, with the exception that VCs are twice more likely to have an MBA degree. Angels are also older and less likely to be female. Lastly, the two are similar in the extent of mentoring support they provide.

Table 2 describes the 253 seed-stage companies in the 2018-2019 cohort. Pre-program information is from startup applications, first session dossiers, and Internet searches. Post-program funding data are sourced from Pitchbook, Crunchbase, founders, mentors, and news articles. Overall, startups are early-stage, science-based, and run by young first-time founders. Due to the scarcity of data on high-technology startups, it is challenging to assess the representativeness of this sample. Instead, the other compares the characteristics of the CDL sample of startups to several other samples and settings of U.S.-based high-technology startups.

The number of founders (2.6) and employees (4.1) is similar to the 2.6 founders and 3.4 employees found in the sample of seed-stage startups in AngelList (Bernstein *et al.*, 2017), and 2.9 founders in the MIT E-Lab startups (Hsu, 2007). Regarding the development stage, 23% have a prototype when applying to the program, which is close to the 29% of university-based projects in the U.S. (Jensen & Thursby, 2001). For IP appropriation strategy, startups are similar to SBIR ventures that received R&D funding in Gans *et al.* (2002), who report a survey score of 3.5/5 for

Table 1: Summary of Mentors

	Angel Investors N = 44		Venture Capitalists N = 148		Difference in Means
	Mean	Standard Deviation	Mean	Standard Deviation	p-value
Experience					
Former Founder	1.00	0.00	0.49	0.50	0.00
Exited Entrepreneur	0.61	0.49	0.32	0.47	0.00
Executive (e.g., CEO)	0.89	0.32	0.94	0.24	0.24
Technical (e.g, data analyst)	0.27	0.45	0.32	0.47	0.52
Academic (e.g., lecturer)	0.05	0.21	0.06	0.24	0.70
Highest Degree					
Bachelor	0.41	0.50	0.30	0.46	0.17
Master (Excl. MBA)	0.14	0.35	0.11	0.32	0.70
Highest Degree: PhD	0.23	0.42	0.18	0.39	0.51
Major					
STEM	0.61	0.49	0.50	0.50	0.19
Business (Excl. MBA)	0.14	0.35	0.16	0.36	0.76
MBA	0.16	0.37	0.38	0.49	0.01
Demographic					
Female	0.07	0.25	0.22	0.41	0.03
Age	51.59	11.32	46.28	10.84	0.01
Mentoring					
Mentorship Hours Committed	27.91	16.20	22.73	20.64	0.13
Unique Startups Mentored	4.50	3.09	4.07	3.36	0.45

Notes: This table compares the characteristics of angel and VC mentors.

Table 2: Summary Statistics of Startups

$N = 253$	Mean	Median	Standard Deviation	Min	Max
Panel A: Venture Characteristics					
Founding Team Size	2.55	2	1.22	1	8
Firm Size	4.13	3	5.52	0	50
Has Prototype	0.23	0	0.42	0	1
IPR Patent Important	0.71	1	0.46	0	1
Pre-Program Capital ($000s)	514.04	67	1,523.20	0	20,000
Pre-Program Revenue ($000s)	149.78	0	485.06	0	5,250
External Funding ($Million)	1.24	0	4.40	0	39
Valuation ($Million)	3.78	0	12.79	0	132
Panel B: Founder Characteristics					
Num. PhD Founders	1.04	1	1.22	0	5
Has PhD Founder	0.55	1	0.50	0	1
Mean Founder Age	34.40	32	8.77	19	68
Has Founding Exp.	0.41	0	0.49	0	1
Has Startup Work Exp.	0.42	0	0.50	0	1
Has Female Founder	0.26	0	0.44	0	1

Notes: This table describes the characteristics of startups.

the importance of patenting. Following their methodology, I label a binary variable from venture applications where founders describe how they intend to protect their IP. 71% state patenting, thought a much lower fraction will file for or be granted a patent–during the 8-month study period, only 13% did.

The median amount of capital raised and revenues generated before joining the program are $67,000 and zero, respectively, reflecting the early stage of the startups in my sample. The mean capital raised before joining the program is approximately USD$370 thousand, which is similar to USD$304 thousand in AngelList startups (Bernstein *et al.*, 2017). Assuming startups were worth close to zero before joining SEP, the three-year step-up in valuation is $2.5 million, which is higher than $2.24 million step-up over eight years in startups that received their first round of VC funding between 2002 and 2010 (Ewens *et al.*, 2018).

Moving to founder characteristics in Panel B, founders are more educated, younger, and less experienced than comparable samples. Half of the teams have at least one PhD founder, twice the startups in MIT E-Lab and MIT Venture Mentoring Services (Scott *et al.*, 2020). The average team age of 34 is lower than the age of 40

found in Ewens *et al.* (2018) and the 2010 Global Entrepreneurship Monitor (Liang *et al.*, 2018), though neither of these samples is constrained to younger seed-stage companies. In terms of experience, 41% of teams have former founders, slightly less than in Ewens *et al.* (2018). Only 26% have at least one female founder, reflecting the documented under-representation of women in tech entrepreneurship (Ruef *et al.*, 2003; Harrison & Mason, 2007).

4 How to Access CDL Data

Interested researchers may access de-identified CDL data upon request and approval by CDL's Chief Data Scientist, and conditional on documented Ethics Review Board approval from the home institution of the Principal Investigator. The application process entails a research proposal submission, an Ethics approval from the home institution of the Principal Investigator, a Data Transfer Agreement (between the PI and CDL), and Non-Disclosure Agreements (for all members of the research team accessing the data). De-identified datasets are then shared via secure file transfer. Datasets are to be deleted permanently from non-CDL servers upon completion of the project, publication of work, or expiration of their ethics approval. Participants consent to this data sharing via the Participation Agreement, and the use of data for such purposes is outlined on the CDL Privacy Policy found at https://creativedestructionlab.com/privacy-policy/.

A number of CDL-connected researchers are already using this data. They already follow this process. We now aim for CDL data to be used by a wide variety of researchers, while preserving the privacy of CDL participants and ensuring the data is used according to the description in the Participation Agreement.

References

Bernstein, Shai, Korteweg, Arthur, & Laws, Kevin. 2017. Attracting Early-Stage Investors: Evidence from a Randomized Field Experiment. *The Journal of finance (New York)*, **72**(2), 509–538.

Ewens, Michael, Nanda, Ramana, & Rhodes-Kropf, Matthew. 2018. Cost of experimentation and the evolution of venture capital. *Journal of financial economics*, **128**(3), 422–442.

Gans, Joshua S., Hsu, David H., & Stern, Scott. 2002. When Does Start-Up Innovation Spur the Gale of Creative Destruction? *The Rand journal of economics*, **33**(4), 571–586.

Harrison, Richard T., & Mason, Colin M. 2007. Does Gender Matter? Women Business Angels and the Supply of Entrepreneurial Finance. *Entrepreneurship theory and practice*, **31**(3), 445–472.

Hsu, David H. 2007. Experienced entrepreneurial founders, organizational capital, and venture capital funding. *Research policy*, **36**(5), 722–741.

Jensen, Richard, & Thursby, Marie. 2001. Proofs and Prototypes for Sale: The Licensing of University Inventions. *The American economic review*, **91**(1), 240–259.

Liang, James, Wang, Hui, & Lazear, Edward P. 2018. Demographics and Entrepreneurship. *The Journal of political economy*, **126**(S1), S140–S196.

Ruef, Martin, Aldrich, Howard E., & Carter, Nancy M. 2003. The Structure of Founding Teams: Homophily, Strong Ties, and Isolation among U.S. Entrepreneurs. *American sociological review*, **68**(2), 195–.

Sariri, Amir. 2024. The Foundations of Advice: A Quantitative Exploration. *Purdue University Working Paper*.

Scott, Erin L., Shu, Pian, & Lubynsky, Roman M. 2020. Entrepreneurial Uncertainty and Expert Evaluation: An Empirical Analysis. *Management Science*, **66**(3), 1278–1299.