

MANUAL FOR DATA LICENSING

YOU PROBABLY SHOULDN'T LICENSE DATA. BUT IF YOU DO — HERE ARE SOME TIPS.

Brickroad builds agentic infrastructure for inference-time private data retrieval. We believe that efficient access to information will change the world.

[DISCOVER](#)[ACCESS](#)[COMPOSE](#)[ROUTE](#)

— INTRODUCTION

HI, MY NAME IS FREEMAN.

[@FREEMAN_LEWIN](#) on TWITTER

- \$100M+ in data & IP licensing negotiated.
- Bootstrapped Brickroad by brokering data deals.
- Should have been a pro-snowboarder.

Read Our Research →



WHERE THIS COMES FROM

Brickroad started by brokering data deals. Now we help labs and financial firms discover, access, evaluate, and route data faster. The “wisdom” in the following slides is based on our experience on both sides of the aisle.

— WHY ARE WE HERE?

EVERYONE IS ABOUT TO EXPERIMENT WITH LICENSING DATA.

We're in an "AdSense" moment. Every single company is starting to think about licensing their data in some capacity. For good reason?

\$32B–\$40B

DATA TRADED TODAY

Strict "data actually sold" basis, across AI, corporate, finance and science.

\$73B–\$99B

PROJECTED BY 2031

AI labs & agents are the fastest growing segment: \$3.9B today → \$34B–\$44B.

100%

BILATERAL, TODAY

Every data deal is negotiated through human one-to-one relationships. It's wildly inefficient.

Labs are flooded

300–400 inbound data pitches a day at a frontier lab. How will you stand out?

Annotators are Buying

Micro1, Mercor, Scale, Handshake are all cold-calling for internal data. Is it worth it?

Buyers are Technical

The counterparty is now an engineer who wants data on demand, at a price, at a quality bar.

Nothing Sells Itself

Marketplaces never cleared. Catalogs don't clear. Someone still has to prove the value.

SIZE YOUR MARKET

Before investing in data licensing – figure out the size of the opportunity.

WHAT'S FOR SALE, AND QUESTIONS YOU SHOULD ASK YOURSELF.

BIZ OPS ARCHIVE DATA

ASK YOURSELF

You license your business

- Would I hand this to a competitor for the same cheque?
- What does my moat look like the year after this sits inside a model?
- Can I anonymize it myself before it leaves the building?

CATALOG DATA

ASK YOURSELF

You license prior work

- Will I ever touch this again — is it genuinely dead?
- Who else has rights in it — clients, collaborators, ex-employees, creditors?
- Am I fine with a one-off price? There is no second deal here.

DISTRIBUTION

ASK YOURSELF

You license your customers

- Did my users consent — and would they say so if asked?
- Am I selling their data, or their attention and labor? The second has a different buyer base.
- Can I route them per task instead of selling them once?

DATA AS EXHAUST

THE ONE THAT CAN BECOME A BUSINESS

The holy grail

- Does it grow every day without anyone doing extra work?
- Is it a time series a buyer would pay for repeatedly?
- Do I own it cleanly — and do my ToS say so?

EVERYONE **OVERESTIMATES** THE VALUE OF THEIR DATA.

Finance's "tenth of a tenth" rule.

\$1,000,000

WHAT THE HEDGE FUND MAKES

A fund runs a trade, or a lab ships a capability, on top of ten combined datasets.



\$100,000

YOUR DATA'S SHARE

Your feed is one input of ten. The buyer can't attribute more than that to you - and won't.



\$10,000

WHAT YOU GET PAID

The buyer keeps most of the value it creates. You get a tenth of the tenth.

MARKET BENCHMARKS

~\$30K

Average hedge-fund data deal. Often \$30K to get done in 30 days.

~\$10K

What some frontier labs will pay for "any data you have." No press release.

\$10K → \$30M

Range of data deals with labs. The distribution is brutal.

~\$300K

Typical one-time annotator offer for an internal archive. So we hear.

BOTTOM LINE

Invest only if it's worth it at these numbers.

EVERYONE UNDERESTIMATES HOW LONG IT TAKES.

Selling data takes time and effort. Finance-firm deals can take over 12 months; AI labs a little less.

MONTH 0-1

MONTH 1-3

MONTH 3-6

MONTH 6-12

Package

Figure out what you actually have. Provenance, rights, a sample, a spec.

Prove

Backtest for funds, evals for labs.
\$1K-\$50K per good backtest;
\$25K-\$100K to set up historically.

Clear

DDQs, security review, redlines.
Enterprise legal alone runs 4-8 weeks. Compliance can sit for six.

Repeat

Non-exclusive is the only realistic route — so you need 10-40 buyers. Then do it again.

12 months

Minimum horizon before licensing is worth judging (finance).

10-40 deals

To make the program matter on a P&L.

3-6 months

Typical enterprise procurement cycle, per deal.

Years

For the multi-million, multi-year contracts we hear about.

IT'S A **DISTRACTION** FROM THE ONLY THING THAT MATTERS.

One hard dependency: you must already be producing first-class data as exhaust from your core operation.

THE SPIEL I GIVE MOST STARTUPS

Make the product great.
Sell the exhaust later.

- You don't have revenue and you're selling the only thing you do all day.
- Every license of your expertise eats at your moat. Non-exclusive deals compound that.
- Labs assume they can reproduce a specialized dataset once they understand it — so they price it low.
- Right sequence: train your own model → make your people faster → sell that product → use licensing as collateral.

THE LOGO TRAP

“I just want [Insert Lab] on the sales sheet for the next raise.”

- 01 Labs will talk to you! Talking to you is a role!
- 02 It doesn't mean you'll get a deal. Most deals are small
- 03 You are not guaranteed a second deal. The logo is the payoff.

Logo validation is real. Just know that's what you bought — and what it cost.

WARNING

Never trust a broker who promises \$\$\$ without effort.

SO WHEN **SHOULD** YOU LICENSE DATA?

If you're still not persuaded, or on the fence – consider whether these apply.

- | | | |
|----|-----------------------------|---|
| 01 | It's exhaust | Falls out of your core operation whether or not anyone buys it. Zero marginal effort to produce. |
| 02 | It's a growing time series | Buyers pay repeatedly for a live feed. Snapshots sell once, at archive prices. |
| 03 | You have a 12-month horizon | And the appetite to close 10–40 non-exclusive deals. You need to invest upfront. |
| 04 | You want revenue, not logos | Want a logo for the next raise? Buy it knowingly — ~\$10K and a release. It is not a business line. |
| 05 | You know who owns it | Clean title, consented users, ToS permitting licensing and training. Creditors cloud it. |

BEFORE YOU LICENSE, CONSIDER

Three things usually worth more than the data

- Train your own model on it and sell the product.
- Route your users, not your rows. Verified, consented, task-ready distribution is what annotators are desperate for — and it pays per task, forever.
- License the dusty catalog you'll never touch again. Cheap, one-off, no regrets.

KNOW WHAT YOU HAVE BEFORE ANYONE ASKS.

Provenance, rights, coverage. Understand what you have now – do an inventory check before a buyer reaches out.

PROVENANCE

What I have

- What does one row represent? One user, one event, one filing?
- Which timestamps does it carry — event time, ingest time, both?
- Is history ever rewritten? Point-in-time correctness is the first thing a fund checks.
- How is it collected, and what breaks the collection?

RIGHTS

Whether I can sell it

- Who owns the data — you, your customers, your users, a creditor?
- Do your ToS and privacy policy permit licensing and ML training? Buyers now watch ToS changes as a signal.
- What is consented, what is contractually restricted, what is PII?
- What must be de-identified — and can you do that yourself, in-house?

COVERAGE

Whether it is valuable

- Universe: tickers, GICS sectors, exchanges, geographies, entity IDs.
- History depth and refresh cadence. Funds want years for backtests; a six-month-old feed is a thin one.
- Volume, growth rate, and what a full year of the feed looks like.
- The one-paragraph description a buyer's agent can match against a query.

MOST COMMON FAILURE

Saying yes to an inbound before knowing which subset of the data is valuable — or legal to sell.

POSITIONING IS THE WHOLE GAME. IT'S SALES!

Data does not sell itself. Funds and labs buy different proof.

FUNDS & QUANTS

Prove the signal

- They buy low-latency signals, not training data. Frame the feed as an input to a trade.
- Run the backtest before they ask. Point-in-time, no lookahead, benchmarked.
- Lead with history depth and refresh cadence. Show the time series growing.

AI LABS

Prove the model impact

- Demand has moved: mass corpora → specific data → production metadata and post-training feeds.
- Quantify lift on a task, compute saved, or an eval they care about — or they assume they can reproduce you.
- Position as an ongoing post-training feed, not a pre-training one-off.

READINESS LADDER

DRAFT

You have a description and a contact.



PRESENTABLE

A buyer reads a sample and a spec without emailing you.



MARKET READY

A buyer can evaluate you without a call.

REMEMBER

“Market Ready” means a buyer can evaluate you without scheduling a call. People don’t like calls.

STRUCTURE THE DEAL SO THERE'S A SECOND ONE.

Most first-time sellers lose on terms they didn't know were negotiable. Get your own counsel — this is not legal advice.

01 Non-exclusive, always

Exclusive won't yield fair value. Non-exclusive sells 10–40 times — the only math that works.

02 Direct beats intermediary

Go straight to the lab or fund. Intermediaries resell to multiple buyers and keep the spread.

03 Scope the release

Past use only. No future scraping, no “all data generated.” Name models, term, purpose.

04 Anonymize in-house

De-identify yourself, before delivery. Don't accept “we'll run it through our pipeline.”

05 Charge for the eval

Backtests and impact studies cost \$1K–\$50K. If they make you provide for them, charge for it.

06 Price the feed, not the file

Consumption or subscription for a live feed. A one-time archive price should be 10× lower.

07 Decide who holds title

Are you the counterparty, or is the data packaged with other sources? Know which.

08 Retention & audit

Term end: deletion, retained records, derived models? Say it in the contract.

09 Plan the sequence

Selling the archive and licensing the feed aren't exclusive. Don't let deal #1 foreclose deal #2.

NOT LEGAL ADVICE

We are not a law firm. But we've seen enough data deals to know where they go wrong — happy to help.

PREPARE YOUR PACKAGE – HERE’S A CHECKLIST.

Built with labs and fund data-strategy teams tired of incomplete inbound pitches. Fill it once, reuse it for every buyer.

01

Company & contact

A reachable human beats a deck. Who signs, who answers technical questions.

02

Description & sample

A snapshot a buyer can diligence without a call. Real rows, schema, gaps.

03

Provenance & integrity

What one record represents, its timestamps, whether history is rewritten.

04

Coverage & Evals

Universe, GICS, exchanges, sample tickers, geographies, eval results and shortfalls.

05

License terms

Approximate a programmable data license, or upload the terms you already use.

READINESS SCORE

Scored as you fill it in. Output: a ZIP any buyer can read, or a listing buyer agents find.

DRAFT

PRESENTABLE

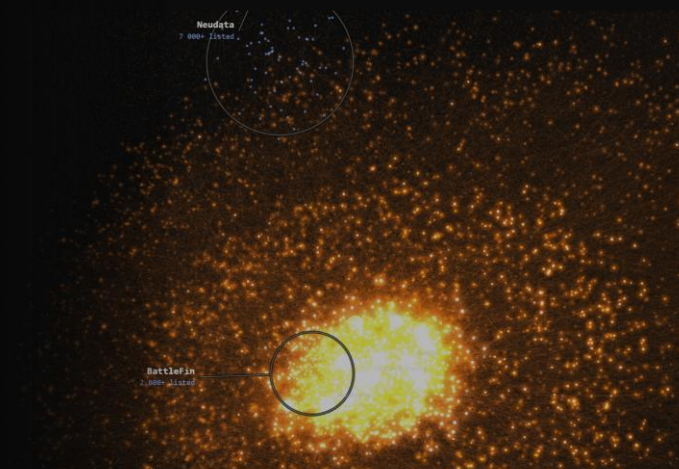
MARKET READY

brickroad.network/supplier-onboarding →

WHY IT MATTERS

Labs see 300–400 pitches a day. The complete, structured package is the one that gets read.

THE MARKET IS MOVING FROM NEGOTIATED DEALS TO **ROUTED FEEDS**.



TODAY

N × M bilateral

- ~\$50K–\$2M all-in per bilateral deal
- ≈3 months from first call to a sample
- Every buyer repeats the same diligence



WITH BRICKROAD - SUPPLIERS CONNECT ONCE

N + M layer

- Package once; connect once (S3, Snowflake, API)
- Agents surface you to matching buyer queries
- Evals and backtests run automatically on samples



ENDGAME

A sustainable data economy

- Company data turns into financial assets.
- Buyers' agents pull just-in-time context at runtime
- Small, frequent, repeatable data transactions.

17,033

suppliers in the registry since May;
88.7% in none of 14 catalogs

38 sec

median time to a verified live sample,
vs ~3 months by vendor trial

35,000+

API endpoints under programmatic
discovery

FOR SELLERS

As AI executes economically critical tasks, data needs to be accessible at the speed of compute. Are you ready?

TL;DR

DATA DOESN'T SELL ITSELF. MOST OF YOU SHOULDN'T TRY. THE REST SHOULD GET MARKET READY.

WHY NOT

Tenth of a tenth · 12 months · distraction
· the Micro1 effect

WHEN YES

Exhaust · time series · 12-month horizon
· revenue not logos · clean title

HOW

Know what you have · position for the
buyer · structure for deal #2

TOOLS

brickroad.network/suppliers — the
checklist, free · happy to read a term
sheet