MAIN STREET WEALTH

A MAIN STREET WEALTH GUIDE FOR BUSINESS OWNERS

THE DATA DIVIDEND

How established companies are getting paid six and seven figures to license the records they already own.

- **The Supply Shift:** why frontier AI labs are licensing private business records right now
- **Valuation & Benchmarks:** verified transactions, payout bands, and the five price drivers
- **Confidentiality Architecture:** five-layer de-identification and what never leaves your control
- **Buyer Due Diligence:** a leading aggregator's pitch examined claim by claim

Sukhrobjon "Rob" Ismoilov, JD, LLM
President, Main Street Wealth LLC

November 2026 Edition
mainstreetwealth.ai

EXECUTIVE OVERVIEW

Table of Contents

01	Beyond the Public Web: The AI Supply Crunch	4
02	Decision Logic: What AI Labs Actually License	5
03	The Market Forces Driving Buyer Demand	6
04	Public Precedents & Landmark Transactions	7
05	Commercial Licensing: Zero Dilution, Zero Debt	8
06	Qualifying Records & Industry Thresholds	9
07	Valuation Drivers & Benchmark Payout Bands	10
08	Deal Math, Net Proceeds & Buyer Rate Cards	12
09	The Five-Layer Confidentiality Shield	14
10	Straight Answers to Owner Objections	15
11	From Assessment to Wire: The 60-Day Roadmap	16
12	The Buyer Ecosystem & First-Mover Window	17
13	Seven Red Flags: Screening Out Bad Actors	18
14	Deconstructing a Buyer's Pitch, Claim by Claim	19
15	Seller-Side Representation: Our Advisory Mandate	20
	Owner & Advisor Q&A	21
	Research, Legal Precedents & Sources	23
	Referral Partnerships & Introductions	24
	Next Step: Request Your Valuation Review	25

ADVISORY PERSPECTIVE

From the President's Desk

I spend my days inside business finances. As an M&A advisor, I see the bank statements, the tax returns, the debt schedules, and the records owners have spent decades building. So when a new category shows up that pays established companies real money for something they already own, with no loan to repay and no ownership to give up, I pay attention.

That is what this guide is. AI companies have run out of the easy data, and they are now paying for the hard kind: the records of how real businesses actually operate. Most owners have never heard of this, and most of what circulates about it is either hype or fear. Neither helps you make a decision.

Main Street Wealth has spent years preparing companies for high-value exits. We now apply the same disciplined, seller-first approach to a new opportunity: licensing a cleaned copy of your operating records to AI companies that need real decision logic. You keep the company. You keep the originals. You keep running the business. You get paid for a processed copy of work you already did.

So I put the whole picture in one place. What the buyers are buying and why. What deals actually look like, with real published numbers. How your records are cleaned and protected before anything is used. The risks, stated plainly, including the parts most promoters skip. And exactly how to find out what your company could be paid.

Read it in one sitting if you can. By the end, you will know more about this market than most of the people trying to sell you something in it.

Sukhrobjon "Rob" Ismoilov, JD, LLM

President, Main Street Wealth LLC

CHAPTER 01

Beyond the Public Web: The AI Supply Crunch

01

Every frontier AI model is built from three core inputs: compute infrastructure, model architecture, and human-generated data. The first two scale with capital. The third has hit a structural bottleneck.

For a decade, AI labs trained on the open internet—books, articles, forums, documentation, and public code repositories. Anything accessible to a web crawler was harvested. That era has closed for three documented reasons.

First, high-quality public text is nearing exhaustion. Quantitative forecasting by *Epoch AI* (Villalobos et al., ICML 2024) estimates the usable stock of public human text at roughly 300 trillion tokens, projecting full exhaustion within the 2026–2032 window. Meanwhile, the *Data Provenance Initiative* at MIT and Harvard (Longpre et al., 2024) documented that over 28% of critical web sources restricted AI crawlers within a single year.

Second, synthetic data cannot replace real human work. A landmark study in *Nature* (Shumailov et al., July 2024) demonstrated that training AI models recursively on AI-generated outputs causes irreversible "model collapse"—erasing rare edge cases and real-world judgment. Models require authentic human workflows to improve.

Third, federal courts priced unconsented scraping in the billions. Copyright rulings in *Thomson Reuters v. ROSS Intelligence* (D. Del. 2025) and the $1.5 billion settlement in *Bartz v. Anthropic* (N.D. Cal. 2025) established a clear rule across corporate legal departments: scraped data is a balance-sheet liability, while consented, licensed data is an institutional asset.

An AI that has read every textbook still does not know how an operating company runs—how a custom quote gets built, how a billing dispute gets untangled, or how a shipment gets rescued at 4 PM on a Friday. That knowledge lives inside private businesses, in their emails, chats, tickets, and SOPs.

THE MARKET WINDOW IN PLAIN TERMS

Frontier models are starving for authentic workflow data, synthetic substitutes degrade model quality, federal courts penalized unconsented scraping, and multibillion-dollar licensing budgets are active today. **The only missing link is business owners who realize their archives are an asset—before their industry category fills.**

CHAPTER 02

Decision Logic: What AI Labs Actually License

02

Let us clear up the biggest misunderstanding first. Buyers are not purchasing your customer list.

They are not buying your client files, your pricing secrets, or a map of your business they can compete against. If a buyer asks for any of that, you are talking to the wrong buyer, and Chapter 13 covers how to spot that.

What they are buying is **decision logic**: the recorded pattern of how experienced people solve real problems at work.

When a duplicate invoice gets caught, escalated, refunded, and patched so it never happens again, that chain of messages and tickets is a lesson in how a business thinks. Multiply that by every problem your team has solved across years of email, chat, tickets, and documents, and you have what the industry calls an **operational data corpus**: the working memory of a company.

Systems That Hold Operational Memory

- **Email threads** (the largest single source in most companies)
- **Team chat:** Slack, Microsoft Teams
- **Documents and playbooks:** Google Drive, SharePoint, SOPs, proposals
- **CRM and sales records:** how deals are worked, won, and lost
- **Support conversations:** how customer problems actually get resolved
- **Code and commit history,** for software companies
- **Tickets and project systems:** Jira, Asana, Confluence, help desks, ServiceTitan, Clio, Toast, and industry platforms

The invoice example, before and after. A real record might read: a named employee at a named logistics company spots a customer billed twice for $12,400, a colleague checks the billing log, finance reissues a refund, a developer patches the script, and the ticket closes in 26 hours. Before a buyer ever sees it, every identifier is transformed. The employee becomes a different name. The company becomes a different company. The amounts change. What survives is the shape of the problem and the steps that solved it. **That shape is the product.** The names were never the value, which is why buyers accept, and insist on, removing them.

RAW INTERNAL RECORD (PRE-PROCESSING)

Maria Lopez at **Acme Logistics** spots that customer **Hartline Manufacturing** was billed twice for the same shipment: **$12,400** in **March**. **Jordan** checks the billing log, confirms the duplicate, and escalates to **Priya** in finance. **Priya** voids the charge and reissues a corrected invoice. A developer patches the billing script. Ticket closed in 26 hours.

→

DE-IDENTIFIED TRAINING RECORD (DELIVERED)

Jennifer at **Northgate Freight** spots that customer **Beacon Industrial** was billed twice for the same shipment: **$9,850** in **June**. **Alex** checks the billing log, confirms the duplicate, and escalates to **Sam** in finance. **Sam** voids the charge and reissues a corrected invoice. A developer patches the billing script. Ticket closed in 26 hours.

Every name, company, and figure is transformed. The problem, the steps, and the judgment survive intact, because that is what was purchased.

This distinction runs through the entire guide: **they want the logic, not the names.** Once you understand that, the safety model in Chapter 9 and the pricing in Chapter 7 both make sense.

CHAPTER 03

The Market Forces Driving Buyer Demand 03

It helps to see the economic forces on the buyers' side, because they explain why capital is committed and why deals are closing now.

The enterprise agent race. A small group of frontier AI labs and hyperscalers are competing to deploy AI agents capable of multi-step knowledge work alongside human teams: drafting, researching, reconciling, coordinating, and resolving exceptions. Annual training and data acquisition budgets at leading labs exceed $1 billion. Operational workflow data is a dedicated line item inside those budgets.

Labor economics vs. software spend. Across the global economy, companies spend roughly six dollars on payroll for every dollar spent on software. An AI system that can reliably assist with real operational tasks addresses a market six times larger than traditional SaaS—but it is useless if it only knows textbook theory. To function like an experienced colleague, it must train on the recorded problem-solving of real teams.

Provenance and auditability. Following federal copyright rulings and regulatory scrutiny, enterprise AI buyers require a verifiable chain of title. A clean commercial license signed by the rightful business owner is no longer optional—it is the legal prerequisite that institutional buyers are paying a premium to secure.

Category buildouts are underway. Airlines, financial publishers, code networks, and thousands of private operating companies are already licensing archives. The training corpora assembled during this cycle will power industry software for the next decade. The practical decision for an owner is whether your company gets compensated for its operational footprint while your sector's quota is open.

> **A COMMERCIAL LICENSE ON SELLER TERMS**
>
> Contrast two eras of technology. Over the past decade, web platforms harvested public content quietly and monetized it without compensating creators. The private operational data market works differently: **defined scope, audited de-identification, seller sign-off, and a direct wire payment.**

CHAPTER 04

Public Precedents & Landmark Transactions 04

You do not have to take any intermediary's word for the scale of this market. The benchmark transactions are documented in SEC filings, court dockets, and corporate disclosures.

Reading the primary filings matters, because press headlines often blur the structure—and the structure shows why private operating archives command strong pricing.

$10M
SPIRIT / GOOGLE · 2026

Spirit Airlines Bankruptcy Auction (Google). During Spirit Airlines' Chapter 11 proceedings, its internal operational archive—over 100 million emails, hundreds of millions of internal messages, software code, and dispatch manuals—was auctioned separately from aircraft or routes. Google won the court-supervised auction with a $10 million bid for the records alone, proving that multi-year operational workflows carry standalone eight-figure value even when detached from physical assets.

$203M
REDDIT SEC S-1 · 2024

Reddit IPO Registration Statement & Google Deal. In its February 2024 Form S-1 filing with the SEC, Reddit formally disclosed $203 million in aggregate multi-year AI data licensing contracts, anchored by an annualized ~$60 million partnership with Google. Human troubleshooting and discussion threads were priced as high-margin institutional assets.

$250M+
PUBLISHER ARCHIVES · 2024

News Corp, Dotdash Meredith (IAC) & Financial Times. OpenAI executed multi-year archive licensing agreements with News Corp (valued up to $250M over five years), IAC's Dotdash Meredith ($16M+ annual baseline disclosed in May 2024 SEC filings), and the Financial Times ($5M–$10M/year) to access structured, fact-checked domain archives.

$1.5B
FEDERAL RULINGS · 2025

Bartz v. Anthropic ($1.5B Settlement) & Thomson Reuters v. ROSS. In 2025, Anthropic agreed to a landmark $1.5 billion class settlement (~$3,000 per book across ~500,000 works, No. 3:24-cv-05417, N.D. Cal.) over shadow-library training sets, while Judge Stephanos Bibas rejected fair use in *Thomson Reuters v. ROSS Intelligence* (D. Del. Feb. 2025). Together, these rulings made licensed provenance mandatory.

2024+
TECHNICAL & VISUAL APIS

Stack Overflow (OverflowAPI) & Shutterstock. Stack Overflow partnered with OpenAI and Google Cloud in 2024 to license validated technical problem-solving threads via *OverflowAPI*, building on the recurring contributor-licensing framework pioneered by Shutterstock and enterprise stock archives.

THE COMMON DENOMINATOR ACROSS EVERY DEAL

Structured human problem-solving, clear title, and documented consent. Established private companies hold the largest unindexed supply of that exact asset.

CHAPTER 05

Commercial Licensing: Zero Dilution, Zero Debt

05

Owners evaluating this usually file it mentally under "funding," so let us be exact about what a licensing payment is not.

It is not equity. You sell no ownership. There is no investor, no board seat, no dilution, no say over your company changing hands. Your cap table, if you have one, never moves. For owners planning an eventual exit, that matters: your equity story stays entirely yours.

It is not debt. Nothing is repaid. There is no interest, no factor rate, no daily or weekly debit, no personal guarantee, no UCC filing against your business, no maturity date hanging over your cash flow. If you have spent any time around merchant cash advances, you know exactly how different that sentence is.

It is not a grant or a program with a catch. It is a commercial license: you own an asset (your records), a buyer rents the right to learn from a cleaned copy of it, and you get paid. The same legal family as licensing a photo, a patent, or a song catalog.

Balance-Sheet & Operational Impact

- The money lands as revenue, and you deploy it: pay down expensive debt, fund inventory, hire, build a reserve, or take a distribution.
- It does not consume your borrowing capacity. Your credit lines, your SBA eligibility, and your banking relationships are untouched. (Tell your lender the truth if asked: it is licensing income, documented by the agreement.)
- It does not require the business to perform afterward. There are no covenants, no reporting treadmill, no one auditing your monthly revenue.
- It can only be done for the first time once. A corpus that has been licensed is no longer virgin supply, which is why the timing chapters of this guide keep returning to the word "first."

> **TAX PLANNING BEFORE PROCEEDS ARRIVE**
>
> Licensing income is taxable income, and the treatment of a large one-time payment deserves a conversation with your CPA before the money lands, not after. Chapter 10 returns to this.

CHAPTER 06

Qualifying Records & Industry Thresholds

06

Not every record qualifies, and not every company has enough of the right kind. This chapter is the honest inventory.

High-Value Operational Records

- **Internal communication with substance.** Email and chat where decisions get proposed, debated, and made. The longer and more connected the threads, the better.
- **Work product and process.** SOPs, checklists, playbooks, templates, proposals, scopes, estimates, post-mortems, training material.
- **Ticket and case histories.** Support tickets, project boards, matter files, claims, work orders: a problem, a path, an outcome.
- **Sales and account records.** How opportunities are qualified, priced, negotiated, and serviced over time.
- **Code and technical documentation,** for software and IT companies: repositories, commit discussions, design docs, incident reviews.
- **Financial and operational workflows.** How billing, collections, purchasing, scheduling, and compliance actually get done inside the company.

Excluded or Non-Qualifying Material

- **Pure personal data:** names, contact details, account numbers, credentials. Buyers do not want it and the cleaning process removes or replaces it.
- **Sensitive content flagged by nature:** certain legal, medical, and personal material is excluded outright rather than cleaned.
- **Thin records.** A company whose work happens by phone and handshake, with almost nothing written down, has little corpus to license. The value lives in what is recorded.
- **Records the company does not own the rights to.** Client-owned or platform-owned material cannot be licensed by you, and a legitimate program will say so.

The desk-work test. What matters is not total headcount but how many people do recorded, desk-based work. A 100-person restaurant group might have 25 people in management, finance, HR, and operations whose work lives in systems. A 30-person accounting firm might have nearly all 30 producing dense, valuable records every day. This is why eligibility thresholds differ sharply by industry, and why professional-services firms punch far above their headcount.

SECTOR / OPERATING PROFILE	TYPICAL MINIMUM TEAM SIZE
Law firms	6 or more people
Accounting, finance, medical	10 to 15 or more
Most other established companies	20 to 50 or more, depending on the program

Rough qualification guide based on published program thresholds.

US-based teams with records primarily in English command the strongest pricing. Canada, Europe, and Australia price similarly. Nonprofits with 20 or more full-time staff can qualify as well; a nonprofit is a business operation with records like any other.

CHAPTER 07

Valuation Drivers & Benchmark Payout Bands

07

Two companies with the same headcount can receive offers five times apart. The difference is five drivers.

Understanding them lets you estimate your own position before you ever take a call.

1. Volume. How much recorded work exists. Remote and hybrid companies generate more, because conversations that would happen in a hallway happen in writing instead. More years of systems, more corpus.

2. History. A company with 15 to 20 years of records carries something a 3-year-old company cannot fabricate: how decisions and playbooks evolved, including through hard cycles. Depth of history is one of the strongest price movers in the market.

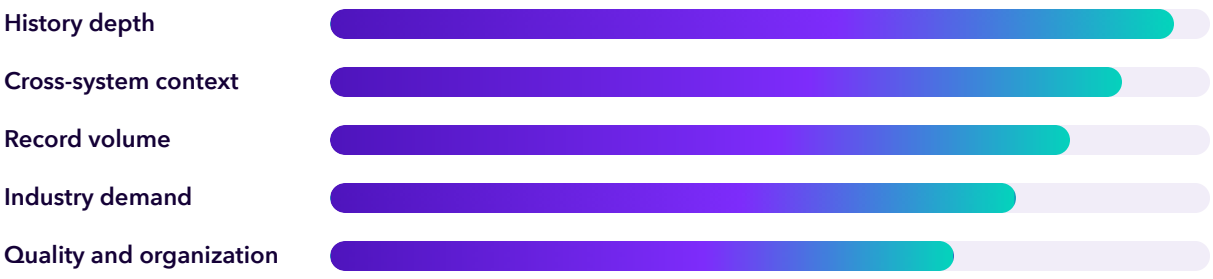
3. Context, the connective tissue. A Slack thread about a product decision, the ticket that implemented it, and the email that announced it are worth far more linked than separate. Buyers are reconstructing how work flows, so records that reference each other across systems raise the value of the whole corpus.

4. Industry demand. Demand rotates with what the labs are building. Legal and accounting have been at the top of the demand list; software and IT, finance, logistics, oil and gas, sales and account management, and customer support are perennially strong. Every category is ultimately case by case, because the buyers are, in effect, recreating the working world industry by industry.

5. Quality and organization. Consistent, well-kept records in standard systems (Google Workspace, Microsoft 365, Slack, Teams, common CRMs and project tools) price higher than the same work scattered across personal accounts and paper.

Primary Corpus Valuation Drivers

What separates a low-band offer from a high-band offer. Relative impact, illustrative.

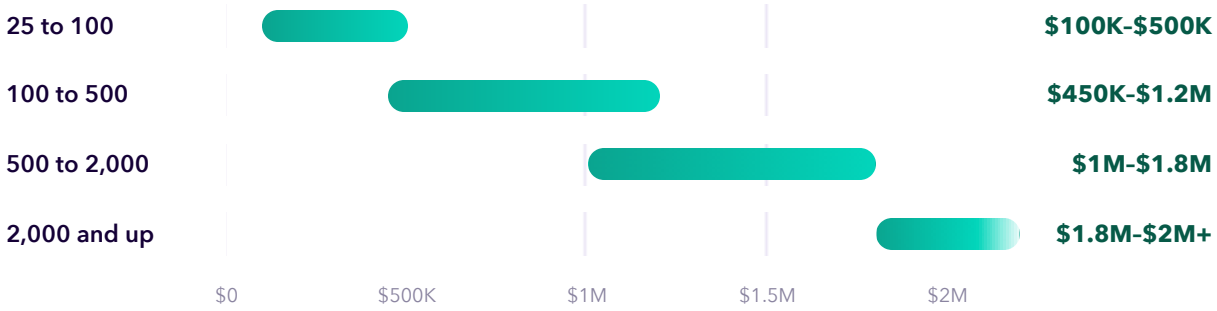


Benchmark Payout Ranges by Headcount

Licensing programs publish outcome ranges. Treat these as the map, not your quote; your number is set only after a sample review of your actual records.

Indicative Licensing Payout Bands

Published ranges of licensing payments by company size (employees). Your figure is set only after a sample review.



COMPANY HEADCOUNT BAND	INDICATIVE LICENSING RANGE
25 to 100	**$100K to $500K**
100 to 500	**$450K to $1.2M**
500 to 2,000	**$1M to $1.8M**
2,000 and up	**$1.8M to $2M and beyond**

Published individual outcomes include **$350,000 to a 60-person accounting firm**, **$450,000 to a 70-person engineering firm**, and **$650,000 to a 100-person logistics company**. These are other companies' results, published by programs active in this market. They are not a promise of your offer, and any honest advisor will tell you the same.

CHAPTER 08

Deal Math, Net Proceeds & Buyer Rate Cards

08

Owners reasonably ask for the math in one place. The structure is simpler than the funding world most owners know.

There is one payment, and it goes to you. The buyer pays a license fee for your corpus after acceptance of the cleaned data. There are no points, no origination fees, no renewal treadmill. If you work with Main Street Wealth, our fee is **15% of the cash you actually receive, and only if you are paid**. It is set out in writing before you sign anything, and if a buyer pays us an introduction fee, that amount is credited against our fee so you are never charged twice.

GROSS LICENSE FEE	ADVISORY FEE (15%)	NET OWNER PROCEEDS	IF NO DEAL CLOSES
$400,000	**$60,000**	**$340,000**	**$0 owed**

Illustration only. Nothing to repay, no equity given up.

Published Buyer Rate Card by Sector (Q4 2026)

These are published buyer ranges for US operating companies. They are advertised estimates, not guaranteed offers. Value depends on usable records, history, quality, and licensing rights.

SECTOR & CORPUS SCOPE	PUBLISHED RATE BAND
Law Firms – Full operating archive (Mode, 20–49 U.S. employees*)	**$200K - $500K**
Accounting & Finance – Full operating archive (Mode, 20–49 U.S. employees*)	**$200K - $500K**
Medical Practices – Administrative records; no health records (Mode, 20–49 U.S. employees*)	**$200K - $500K**
Software Companies – Private legacy code (COBOL, .NET, Java)	**$10K - $100K+**
Other Small Businesses – Company-wide operating history (20–50 employees)	**$10K - $100K**
Logistics & Operations – SOPs, fulfillment and inventory workflows (30+ employees)	**$100K+**
Manufacturing & Engineering – CAD, inspections, maintenance and project records	**Custom Quote**

*Mode: same published size band for all three sectors. Requires 3+ years of records and 5+ systems. Advertised estimates, not guaranteed offers or completed-deal averages.

Illustrative Company Profiles

- **A 40-person professional services firm,** 12 years of records in email, chat, and a practice system, in a high-demand category. Published bands suggest the low-to-mid six figures; companies with this profile have published outcomes around $250K to $450K.
- **A 120-person logistics company,** deep ticket and dispatch history across connected systems. Published outcomes for this profile cluster around $450K to $750K.
- **A 15-person law firm.** Small headcount, extremely dense records, top-of-market demand. Firms like this are exactly why headcount alone does not price a corpus; outcomes in the six figures are published for firms a fraction of the size of the logistics example.

What moves you up your band: more years of history, records that connect across systems, a category in current demand, and clean, well-organized data in standard tools. **What moves you down:** thin written records, short history, fragmented systems, and a category the buyers have already filled.

Speed. Programs quote roughly 60 days from first conversation to payment as a target when a file moves cleanly. Plan your expectations on up to about three months. Either way, measure it against every other six-figure event in business life: no underwriting committee, no closing table, no repayment schedule on the other side.

> **RETIRING HIGH-COST DEBT WITHOUT DILUTION**
>
> Owners who have been through expensive short-term debt will feel the difference immediately. A **$300,000** license payment, after our 15% fee, is **$255,000 you keep**, with nothing to repay. The same $300,000 taken as a typical short-term advance can cost $90,000 to $150,000 in factor charges and leave in weekly debits inside a year. A licensing payment used to retire that kind of debt does double work: it removes the balance and the bleed at the same time.

CHAPTER 09

The Five-Layer Confidentiality Shield

09

This is the chapter the whole decision turns on, so it gets the most detail.

A legitimate licensing program protects you in layers aligned with federal de-identification frameworks (including NIST SP 800-188 and SP 800-122), and you should be able to verify every layer before you agree to anything.

LAYER 1

Seller-defined scope. Before any export, you decide which systems and which date ranges are in, and which are out. Entire departments, date windows, or record types can be excluded. The scope is written into the arrangement before a single file moves.

LAYER 2

Automated PII & entity transformation. The exported records go through a multi-stage de-identification pipeline:

- **Automated detection** of direct and indirect identifiers: names, emails, phone numbers, account numbers, tax IDs, and credentials.
- **Pseudonymization, not crude deletion:** names, company names, and figures are replaced with consistent synthetic tokens, so the decision logic of a thread survives while its identity does not. (As in Chapter 2's invoice example: the employee, company, and dollar amounts all change; the workflow chain stays intact.)
- **Sensitive-content exclusion:** privileged legal, clinical, or restricted personal material is suppressed outright rather than cleaned.
- **Isolated staging environments:** data moves through access-controlled pipelines, never via unsecured attachments or public links.

LAYER 3

Mandatory seller sample audit. Before anything is delivered to a buyer, you review a cleaned sample of your own corpus. If something should not be there, it comes out. Nothing is delivered on the strength of a promise; you inspect the de-identified output first.

LAYER 4

Direct commercial license execution. The legal instrument is a license you execute yourself, with the buyer or program, after the sample audit. It defines what was licensed, in what cleaned form, and for what permitted training use. Until you sign, nothing is licensed.

LAYER 5

Unconditional walk-away rights. At every step before signing, you can stop, and nothing has left your company. Initial valuation ranges are built strictly from high-level metadata (industry, headcount, years in operation, software stack) without touching raw files.

THE NON-NEGOTIABLE SECURITY BOUNDARY

Nobody in a legitimate process ever asks for your client files, your passwords, or your logins. Not the buyer, not the program, not the advisor who introduced you. **If anyone does, end the conversation.** That single rule screens out nearly every bad actor in this market.

CHAPTER 10

Straight Answers to Owner Objections

10

Does licensing our records train AI to automate away our own team?

It trains AI on how work gets done, the same material every training manual, hire, and consultant already carries out of every company. Two honest points. First, the labs are building these systems regardless; the data is being assembled now, from companies that chose to be paid for it. Refusing to license does not slow the technology down; it only decides whether your company shares in the economics. Second, the near-term design goal the buyers describe is agents that work alongside teams on the work nobody wants: reconciliation, drafting, chasing, summarizing. What you are licensing is patterns of problem-solving, not your team's jobs, and the cleaned data contains nothing that identifies your company as the source of any specific practice.

Could a buyer extract our client list or proprietary trade secrets?

The buyers receive de-identified material with names, companies, and figures transformed. A client list would be worthless to them even if it survived cleaning, which it is designed not to, because identities are the one thing with no training value. As for secrets: you control the scope. Genuinely proprietary material, the formula, the book, the bid strategy you would never show a competitor, can be excluded from scope or filtered as sensitive content, and you verify that in the sample audit before signing.

Why can't AI labs simply scrape this material without paying?

Because it is not public. Your internal records sit behind your logins, protected by law and by your ownership. That is precisely why they have value: this is the one major class of training data that cannot be taken, only licensed.

What if our historical systems are unorganized or fragmented?

Messy is normal, and it mostly affects price, not eligibility. Gaps, older systems, and partial history reduce the corpus value; the sample review is what prices reality. What disqualifies a company is not disorder, it is absence: too little recorded work to learn from.

How are licensing proceeds treated for corporate tax purposes?

A licensing payment is income to the business, and a six-figure one-time payment deserves planning. Whether it is ordinary income, how timing interacts with your fiscal year, and how existing losses or carryforwards apply are questions for your CPA with the actual agreement in hand. We are M&A advisors, not your tax advisor, and anyone who answers tax questions glibly is telling you something about themselves.

Will a data license complicate a future M&A exit?

A license of a cleaned copy of your records does not transfer ownership of the company or the originals. Like any material contract, it should be disclosed to a future buyer, and the proceeds simply show up in your financials as licensing income. Because we prepare companies for exits every day, we review license terms with that future sale in mind.

Should our outside corporate counsel review the license agreement?

You will sign a license with real money attached. Having your own attorney review it is money well spent, and a legitimate program will never pressure you to skip that step. The same goes for this guide: it is education, not legal advice.

Is this a recurring revenue stream or a one-time event?

A corpus can be licensed for the first time only once. Programs may license future periods of new records in some cases, but the founding event, the deep historical corpus, happens once. It is one of the two real timing pressures in this market. The other is next.

CHAPTER 11

From Assessment to Wire: The 60-Day Roadmap

11

Owners who have survived commercial underwriting tend to brace for a gauntlet. This process is shorter, and it starts with a conversation, not an application portal.

1

15 MINUTES · NO COST

Confidential Corpus Assessment

A 15-minute call or our short evaluation form: industry, systems, years of records, rough counts, and how many people do desk work. If it is not worth taking to buyers, we tell you in writing, say so plainly, and stop. It has cost you fifteen minutes.

2

INDICATIVE QUOTES · NON-BINDING

Competing Written Valuation Ranges

With your permission, we take your company to licensing programs and bring back two or more written indicative ranges. The program you choose can issue a non-binding letter of intent that holds its range while the review runs (typically a 14-day window). You pick, or you pass. Nothing is exclusive until you decide.

3

SELLER SCOPE · SAMPLE AUDIT

Scoped Export & De-Identification Audit

Under the scope you approved, records are exported and run through the cleaning pipeline described in Chapter 9. You review the cleaned sample and approve it, or send it back.

4

TARGET ~60 DAYS END TO END

License Execution & Wire Settlement

Your counsel reads the license. You approve the categories. The buyer accepts the cleaned corpus, you sign the license, and payment follows on the program's stated terms (net 30 after acceptance is the common structure).

End to end, programs target about 60 days from the first call. Budget up to three months in your own planning and be pleasantly surprised rather than the reverse.

PURE CONTINGENCY ALIGNMENT

Nothing upfront. No retainer, no monthly fee, no charge for the call, no cost if nothing closes. If your company licenses its records, our fee is 15% of the cash you actually receive, disclosed in writing before you sign anything, with any introduction fee a buyer pays us credited back to you.

CHAPTER 12

The Buyer Ecosystem & First-Mover Window

12

The buyers, in plain terms. At the top sit the frontier AI labs and the largest technology companies, the names you already know, competing to build the agents described in Chapter 3. Beneath them sits a growing layer of licensing programs and data companies that source, clean, and package corpora from private businesses for those buyers. When you license through a program, the program is doing the industrial work (sourcing, de-identification, quality review, delivery) and the lab is the end customer. This structure is normal in young markets: the end buyers do not maintain storefronts for ten thousand individual companies, so specialists aggregate the supply.

Why "first" matters. A buyer building the legal category needs the strongest legal corpora it can get, this year. Once a category is covered to the depth the models need, the same records are worth a fraction, because the marginal value of the ten-thousandth similar corpus is small. This is not a sales line; it is how training data economics work, and it is visible in every maturing data category before this one. The window is not "someday prices drop." It is that the founding purchases in each category are the best-priced ones, and they are happening now.

> **WHY FOUNDING CORPORA COMMAND PREMIUMS**
>
> The first companies in a niche with a deep file are getting the quotes. **The tenth will not.**

Why this is a boom and not a bubble to wait out. Booms built on speculation collapse when the story ends. This one is built on an input the buyers cannot manufacture: consented, real-world operational records, with a legal settlement history that made consent mandatory and a public-data supply that is measurably running out. The demand side is funded at the billion-dollar scale and the supply side is a fixed number of established companies. When funded demand meets fixed supply, prices do what prices do.

Where the world is headed. Within a few years, AI agents trained on real operations will be standard equipment in your industry, quoting, scheduling, reconciling, and drafting beside your team. Your competitors will use them whether or not they licensed anything. The owners reading this guide are deciding one narrow thing: whether their company also got paid, once and meaningfully, for helping teach those systems how the work is done.

CHAPTER 13

Seven Red Flags: Screening Out Bad Actors 13

A new money category attracts new predators. This market is no exception. Protect yourself with these seven screening tests, which any legitimate program will pass without hesitation.

WARNING 01

Anyone who asks for raw data, client files, credentials, or logins early. The legitimate sequence is facts first, scope second, cleaned sample third, raw nothing, ever, outside the controlled export you approve. An upfront grab for your systems is the loudest red flag in the market.

WARNING 02

Anyone who asks you to sign exclusivity before an estimate exists. An early, broad exclusive can lock your one first-license event to a party who may never perform. Paper should follow numbers, not precede them.

WARNING 03

Asset sales disguised as licensing. Some offers are purchases of your data outright, or perpetual rights with no defined use. A license has a defined corpus, a defined use, and a defined payment. Read for those three definitions, or have your attorney find them.

WARNING 04

Buyers who cannot explain de-identification mechanics. Ask exactly how names, figures, and sensitive content are handled, and when you see a cleaned sample. Vague answers mean the pipeline is a promise, not a process.

WARNING 05

Guaranteed payout figures before a sample review. Nobody honest can price your corpus from your headcount alone. Ranges, yes. Guarantees, no. This guide gives you ranges and published outcomes for the same reason.

WARNING 06

Manufactured countdown pressure. The timing case in Chapter 12 is real, but it is structural, not a countdown timer. Anyone who needs your signature this week, before your attorney sees paper, is telling you who the deal is designed to protect.

WARNING 07

Upfront fees to "get you listed." Exploring should cost you nothing. Any advisor's compensation should be contingent on you actually being paid, stated as a clear percentage, and disclosed in writing before you sign. An upfront listing fee is not how this market works.

CHAPTER 14

Deconstructing a Buyer's Pitch, Claim by Claim

14

Most owners will never hear a licensing program explain itself at length.

One of the largest programs in this market, Mode Inc, publishes a detailed presentation by its founder and CEO, Dan Novaes, walking through the entire thesis. It is a sales video, and it should be read like one. But it is also the clearest public statement of how the buying side thinks, so here is what he says, broken down, with our notes where the record needs correcting.

"The data wall is real, and consent is the product."

Novaes frames the market exactly as Chapter 1 does: labs consumed the public internet, unconsented data produced lawsuits (he cites the same $1.5 billion Anthropic settlement), and consented private data is now the scarce input. **Our note:** this is the rare sales claim that the public record fully supports.

"We received offers in the range of $6 million for our own data."

Mode is a serial acquirer of apps and businesses. Novaes says the licensing idea began when buyers priced Mode's own operational records in the millions, and that Mode ran the full process, export, de-identification, delivery, on its own portfolio companies before offering it to anyone. **Our note:** ask any program the question he suggests asking. Did you license your own data first? It is a genuinely good test.

"$100,000 to $5 million, and here is what moves the number."

His stated range for company payouts matches the published bands in Chapter 7, and his five drivers are the five in this guide: volume, history, cross-system context, industry demand, and quality. His worked example: a duplicate-invoice incident whose cleaned version teaches the lab nothing about the company and everything about how billing problems get solved. **Our note:** his eligibility floor in the video is looser than the programs' published partner thresholds. Quote yourself against the stricter published numbers in Chapter 6, not the loosest ones in a pitch.

"About 60 days, application to payment."

He lays out the same four steps as Chapter 11: confidential intake, indicative range, a non-binding letter of intent with a 14-day window, export and de-identification, then payment on net-30 terms after lab acceptance. **Our note:** 60 days is the program's stated target. Our standing advice to clients is to plan on up to three months and treat faster as upside.

"Google paid $10 million for Spirit's records."

He uses the Spirit deal as his proof anchor, as does nearly everyone in this market. **Our note, and it matters:** the bankruptcy record shows a $10 million winning bid at auction, subject to court approval, not a completed payment at the time he recorded it. The distinction does not weaken the story. It is the price a court-supervised auction put on a failed company's records. But a market this young runs on trust, and you should notice which advisors bother to get the public facts right.

"Not equity, not debt. Free money for records you already own."

His phrase "free money" is salesmanship; the accurate version is Chapter 5. It is a license payment: real, non-dilutive, non-repayable, and earned by an asset you built over years without knowing it was one.

> **VERIFYING BUYER CLAIMS AGAINST THE PUBLIC RECORD**
>
> The thesis in this guide is not a sales invention. It is the buying side's public position, stated on the record, with numbers attached.

CHAPTER 15

Seller-Side Representation: Our Advisory Mandate

15

We have spent years preparing companies for high-value exits. We bring the same disciplined, seller-first approach to licensing your records.

That means we sit on your side of the table: we scope the file with you, present your company properly, put competing ranges in front of you, read the paper with your counsel, and stay on the file until you are paid, or until you decide to walk away.

MANDATE TERM	HOW MAIN STREET WEALTH REPRESENTS YOU
Who we work for	You. Not the AI company. We do not take a license of your data.
What you pay upfront	Nothing. No retainer, no monthly fee.
Our fee	15% of the cash you actually receive. If a buyer pays us an introduction fee, that amount is credited so you are not charged twice.
If they pass	You pay nothing. You keep every file.
What you give up	A cleaned copy of the categories you approve. You keep the originals and keep running the business.
Your veto	You can decline any buyer, any category, and any draft license. No sample leaves without your written yes.

Four Steps to Initiate a Review

1

Send back the evaluation form, or book the call

Systems, years, rough size, what you already know you will not include. Any industry.

2

We tell you, in writing, whether it is worth taking to buyers

If it is not, we say so and stop.

3

If it is, we get two written ranges

You pick, or you pass.

4

Your counsel reads the license

You approve the categories. Then, and only then, does anyone export.

QUICK REFERENCE

Owner & Advisor Q&A

What exactly is being licensed?

A cleaned, de-identified copy of your company's operational records: the email threads, chats, tickets, documents, and workflows that show how work gets done. Not your client list, not your identity, not your secrets.

Is this a loan? Do I pay it back?

No. It is a license payment. Nothing is repaid, there is no interest, and no lien is placed on your business.

Do I give up any ownership of my company or my data?

No equity changes hands, and you license a cleaned copy of your records; you do not surrender your systems or your rights in your own business information beyond the license you sign and approve.

How much could my company get?

Published program ranges run from about $100K for smaller qualifying companies to $2M and beyond for very large ones, with most established mid-size companies landing in the six figures. Your number is set only after a sample review of your records.

How fast does it happen?

Programs target about 60 days from first call to payment. Plan on up to three months.

What does it cost me to find out?

Nothing. The call is free, the assessment is free, and if nothing closes you have spent a conversation. If a deal closes, our fee is 15% of the cash you actually receive, set out in writing before you sign anything.

Who sees my data?

A controlled pipeline: export under your approved scope, automated de-identification, sensitive-content filtering, and a cleaned sample that you personally review and approve before anything is delivered to a buyer.

Will my clients or employees be identifiable?

The process is designed so they are not: names, companies, and figures are detected and transformed, sensitive material is excluded, and you verify the result in the sample audit.

What if we are small?

Professional-services firms qualify at small headcounts (law firms from about 6 people; accounting, finance, and medical from about 10 to 15) because their records are dense. Most other industries need a larger desk-based team. The assessment exists to answer this honestly for your company.

We are mostly field or floor staff. Does that matter?

What counts is desk-based, recorded work. A large field operation with a real back office can qualify on the back office's records.

Does my industry qualify?

Legal and accounting are in the highest current demand, with software and IT, finance, logistics, oil and gas, sales and account management, and customer support also strong. Every category is ultimately case by case.

Can a nonprofit do this?

Yes, with 20 or more full-time staff. A nonprofit's operations generate the same kind of records.

What if my records are messy or split across systems?

That affects price more than eligibility. The sample review prices what actually exists.

Will this hurt my credit or my ability to borrow?

No. There is no debt, no lien, and no credit inquiry involved in licensing. It is documented licensing income.

Can I consult my attorney and CPA?

You should, and any legitimate program expects it. Bring them the actual agreement before you sign.

Can I change my mind?

At any point before you sign the license, yes. Nothing is licensed, exported for delivery, or used before your signature and sample approval.

Can I license again later?

The first license of your historical corpus is a one-time event. Future periods of new records may be licensable in some cases, which is a conversation for after the first deal.

Why not approach the AI companies myself?

You can try. The frontier labs do not run storefronts for individual mid-size companies, which is why the sourcing and cleaning layer exists. Going through a program also puts the de-identification industrial work on specialists instead of your team, and having an advisor on your side means competing ranges and a second set of eyes on the paper.

Why is Main Street Wealth involved in this at all?

We have spent years preparing companies for high-value exits, which means we sit inside business finances every day and know how to present a company to a sophisticated buyer. This category pays our kind of client for an asset most have never priced. We work for you, not the AI company, and we do not take a license of your data. We scope the file with you, present the company properly, bring you competing ranges, and stay on the file with you. Any offer is made solely by the licensing program after its own review.

PRIMARY SOURCES

Research, Legal Precedents & Sources

The analysis in this guide draws on peer-reviewed machine learning research, federal copyright litigation dockets, SEC registration filings, and federal data-governance standards.

1. Academic & Technical Research on AI Data Supply

- **Villalobos, P., Ho, A., Sevilla, J., Besiroglu, T., Heim, L., & Hobbhahn, M. (2024).** *"Will we run out of data? Limits of LLM scaling based on human-generated data."* Proceedings of the 41st International Conference on Machine Learning (ICML 2024) / Epoch AI (arXiv:2211.04325). Quantifies the stock of high-quality public human text (~300T tokens) and projects exhaustion between 2026 and 2032.
- **Shumailov, I., Shumaylov, Z., Zhao, Y., Papernot, N., Anderson, R., & Gal, Y. (2024).** *"AI models collapse when trained on recursively generated data."* **Nature**, Vol. 631, pp. 755–759 (July 2024). Demonstrates that training models on synthetic AI output causes irreversible "model collapse" and erases tail-distribution decision logic.
- **Longpre, S., Mahari, R., Obeng-Marnu, N., Brannon, W., South, T., Gero, K., Pentland, A., & Kabbara, J. (2024).** *"Consent in Crisis: The Rapid Decline of the AI Data Commons."* Data Provenance Initiative (MIT, Harvard & CMU) / NeurIPS 2024 (arXiv:2407.14933). Longitudinal audit of 14,000 web domains showing over 28% of critical web sources blocked AI crawlers within one year.

2. Federal Court Rulings & Copyright Settlements

- ***Bartz et al. v. Anthropic PBC*, No. 3:24-cv-05417 (N.D. Cal. 2025).** Split summary judgment by Judge William Alsup distinguishing lawful acquisition from shadow-library piracy (LibGen/PiLiMi), followed by the landmark **$1.5 billion class-action settlement** (~$3,000 per work across ~500,000 books) approved by Judge Araceli Martínez-Olguín.
- ***Thomson Reuters Enterprise Centre GmbH v. ROSS Intelligence Inc.*, No. 1:20-cv-00613 (D. Del. Feb. 11, 2025).** Summary judgment opinion by Judge Stephanos Bibas rejecting fair-use defenses where proprietary Westlaw legal headnotes were used without a license to train a competing AI research tool.

3. SEC Filings, Bankruptcy Dockets & Corporate Licensing Disclosures

- **Reddit, Inc. – SEC Form S-1 Registration Statement (Feb. 22, 2024).** U.S. Securities and Exchange Commission filing disclosing $203.0 million in committed multi-year AI data licensing revenue across 2024–2026, including its annualized ~$60 million partnership with Google.
- ***In re Spirit Airlines, Inc.* – Chapter 11 Operational Archive Auction (2026).** Court-supervised bankruptcy auction in which Google submitted the winning $10 million bid for Spirit's standalone archive of 100M+ emails, internal messages, software code, and operational manuals.
- **Enterprise Publisher & Technical API Disclosures (2024–2025): IAC Inc. / Dotdash Meredith** (Form 10-Q & May 2024 investor disclosures detailing an annual baseline of $16M+ from OpenAI licensing); **News Corp** (May 2024 multi-year OpenAI agreement reported at up to $250M); **Financial Times Group** (April 2024 enterprise archive license reported at $5M–$10M/year); and **Stack Overflow** (May 2024 *OverflowAPI* agreements with OpenAI and Google Cloud).

4. De-Identification Standards & Program Benchmarks

- **Garfinkel, S., Near, J., & Dajani, A. (2023).** *NIST Special Publication 800-188: De-Identifying Government Datasets: Techniques and Governance*, and *NIST SP 800-122: Guide to Protecting the Confidentiality of Personally Identifiable Information (PII)*. U.S. National Institute of Standards and Technology frameworks governing pseudonymization, token replacement, and human disclosure review audits.
- **Aggregator & Buyer Program Benchmarks (2025–2026):** Published sector rate cards, eligibility floors, and founder technical briefings from active private-corpus aggregators (including Mode Inc. / Dan Novaes operational data briefings and Q4 2026 buyer rate cards).

METHODOLOGICAL NOTE ON MARKET FIGURES

Figures cited from SEC filings and court dockets reflect official public records. Figures described as "reported" reflect financial-press disclosures of private commercial contracts. Published buyer rate bands are indicative market benchmarks and do not constitute a binding offer or valuation guarantee for any individual company.

PARTNER NETWORK

Referral Partnerships & Introductions

If your own company turns out not to be a fit, this guide can still pay off for you.

Most companies in this market arrive through an introduction. If you know an owner whose company has built up years of operational records, or you are a CPA, financial advisor, or attorney who works with business owners and wants to make introductions yourself, we would like to hear from you.

Ideal Introduction Profiles

FELLOW BUSINESS OWNERS

Established operating companies

Law firms (6+ people), accounting, finance, and medical practices (10–15+ people), and established companies across logistics, software, engineering, and operations (20–50+ people) with years of desk-based work in standard systems.

TRUSTED ADVISORS

CPAs, attorneys, and financial advisors

Advisors who sit inside client finances and want to bring business owners a non-dilutive, debt-free licensing opportunity—or partner with us on introductions across their client base.

Three-Step Partner Workflow

1

DIRECT EMAIL OR PORTAL

Send a short introduction

Send the introduction to rob@mainstreetwealth.ai with the subject line "Data Dividend Introduction," or reach us through mainstreetwealth.ai/contact.

2

SELLER-FIRST HANDLING

Confidential assessment and referral terms

We take it from there, treat the owner with the same confidentiality we give every client, size up the file in fifteen minutes, and walk you through how we work with referral partners.

3

ZERO PRESSURE

The owner decides with full numbers in hand

If the company qualifies, we bring back competing written ranges. The owner and their counsel approve every step before any record is exported or licensed.

OUR COMMITMENT TO EVERY INTRODUCED OWNER

No upfront fees for the owner. Our fee is contingent on the owner actually being paid, and every term is disclosed in writing before anything is signed.

GET STARTED

Next Step: Request Your Valuation Review

You now know how this market works, what buyers pay for, how protection is structured, and what to watch out for. The only question left is the one this guide cannot answer for you: what your company's records are worth.

Finding out costs one short call, and there are two ways to start.

PATHWAY 1

Schedule a 15-Minute Valuation Call

Fifteen minutes. We size up your company together, tell you straight whether you qualify, and if you do, what range to expect and what happens next.

Book at mainstreetwealth.ai/contact

PATHWAY 2

Submit Company Profile via Email

Open a prefilled email, fill in the blanks, and send it. We take it to buyers from there. Or write to rob@mainstreetwealth.ai directly with:

Open the prefilled email

- Your name and title
- Company name
- Company website
- What the company does (industry)
- Number of employees (and how many are desk-based)
- Years in business
- Where the company is based (city, state)
- Where your work records live (email, chat, CRM, industry software)
- Roughly how many years of records you have
- Anything you want to exclude from scope
- Best phone number

Or call +1 917-710-7950.

Your records took years to build. The buyers building the next decade of business tools are paying for exactly that, once, at founding prices, to the companies that move first. Find out what yours are worth while the window described in this guide is the one we are still in.

MAIN STREET WEALTH

Sukhrobjon "Rob" Ismoilov, JD, LLM
President, Main Street Wealth LLC
1123 NE 17th Ave, Fort Lauderdale, Florida 33304
rob@mainstreetwealth.ai · +1 917-710-7950 · mainstreetwealth.ai

MAIN STREET WEALTH

Your records took years to build. Get paid for them once, on your terms.

Main Street Wealth LLC

1123 NE 17th Ave, Fort Lauderdale, Florida 33304

rob@mainstreetwealth.ai · +1 917-710-7950

mainstreetwealth.ai

Main Street Wealth LLC is an independent advisory firm that represents companies in data licensing transactions and is not the buyer of any data. All figures in this guide are indicative ranges, published outcomes from other companies, advertised estimates, or reported transactions; they are not a guarantee, quote, or promise of any offer. Any offer is made solely by a licensing program after its own review. Nothing is exclusive until you sign the advisory agreement, and nothing is licensed until you sign a license. This guide is education and an offer of representation, not legal, tax, or financial advice. Review any agreement with your own attorney and tax advisor before signing.