

Correct is Not Enough

*Why meaning, not models, decides
the future of investment research*

White paper



Executive Summary

Anyone who has worked at a research desk knows that adjusted earnings, credit spreads, and private-market returns can mean different things at different firms. For a long time that worked well. A trained reader spotted the difference, adjusted the model, and left a footnote for whoever picked it up next.

A research report has two jobs. It has to be right, and it has to be understood the way the analyst meant it. Our industry has spent 30 years and a great deal of money on the first job and almost nothing on the second. That worked while every report ended up in front of an experienced human being. It works much less well when the reader is a model, a data pipeline, a client's own AI, or an analyst with nine minutes before a call. None of them reads footnotes.

THE SHORT VERSION

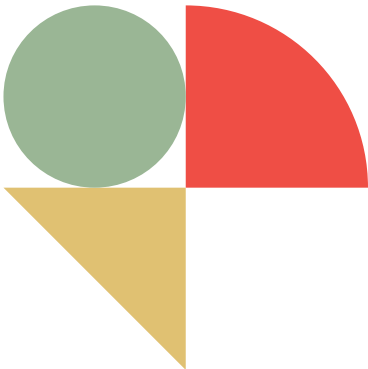
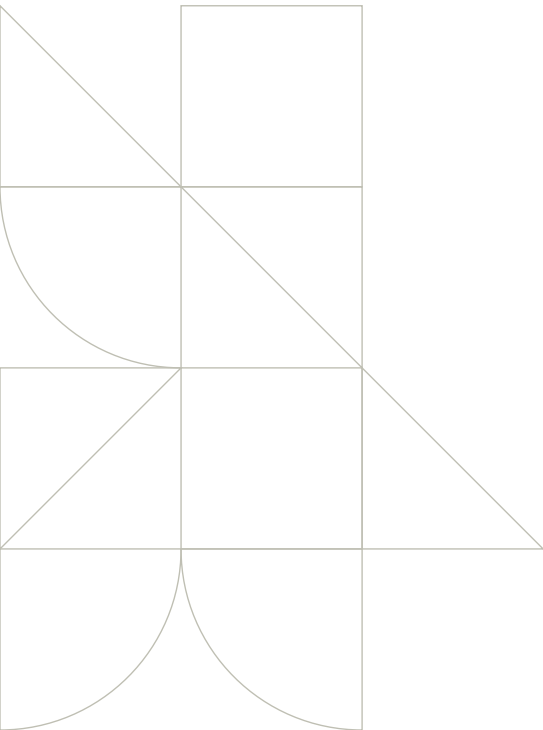
The two jobs multiply, they do not add. A report that is right but misread carries a wrong conclusion with the authority of a right one.

The measures that drive investment decisions are the ones the rulebooks leave open. Regulators govern how adjusted earnings and free cash flow are presented, not how they are calculated.

Somebody has to write down what each measure counts, what it leaves out, and which comparisons hold. Almost no firm does. A knowledge graph is the most direct place to keep that record, though not the only one.

Settle the meaning first and publishing becomes a formatting problem. Note, summary, article, feed, API, podcast, and voice all inherit it.

Ninety days on one contested measure will tell you whether this works at your firm. It comes out of a research budget, not a capital plan.



Being Right Is Only Half the Job

1. The two jobs of a research report

An analyst compares two companies on adjusted earnings. Both sets of accounts are audited. Both disclosures follow the rules on non-standard measures to the letter. One company adds back stock compensation and the other does not. The analyst writes up a margin gap of a few hundred basis points, and roughly half of it is not performance at all. It is arithmetic.

Nothing failed here. Everything was disclosed, and a reader who went looking would have caught it, which is precisely why nobody fixed it. But a footnote was never documentation. It was one professional writing to another in shorthand, assuming judgment at the other end. Take the human out of the loop and the shorthand stops meaning anything.

Two decades of spending on data quality answered one question well: Is this number right? It never asked the harder one. Right about what?

Call the gap between what a firm’s numbers say and what they mean semantic debt. Like technical debt, it built up quietly because people kept paying the interest, one reconciliation by hand at a time, one convention passed on by sitting nearby. Nobody paid down the principal.

Listing troublesome measures does not help, because that list has no end. What helps is sorting the ways a number can mislead you.

EXHIBIT 1: Seven ways a number misleads

Type	What goes wrong	What it looks like
Things companies do—you have to model these whether you like it or not		
1. Same label, different math	Two companies compute the same measure differently	One adds back stock compensation; the other does not
2. Different rulebook	The same idea measured under a different accounting standard	IFRS 16 pulls lease cost out of EBITDA; a US peer keeps it in
3. The label stays, the contents change	A company changes what its own measure covers between periods	Coinbase revised what its Adjusted EBITDA excludes in Q1 2024 and recast prior periods to match
4. A different slice	The same label covering a different part of the business	Whole company against one segment; gross against net; restated against as first filed
Things your firm decides—or has not decided yet		
5. Nobody owns it	No one is accountable for the house view	“Recurring” settled two ways by two analysts on one sector
6. Two official answers	Two systems each publish an authoritative number and they do not match	Risk and research disagree on leverage for the same issuer
7. An undated change	A convention changes and nobody records when	A Q1 note cannot be rebuilt in Q4 on the basis that applied then

Type 3 is worth pausing on: Coinbase did the careful thing and restated its earlier figures, and anything already published on the old basis still does not tie. The first four types are facts about the outside world. The last three are choices inside your own house, and each could be settled by a decision.

2. What it costs

Here is the part that surprises people. You do not need two companies to disagree for this to bite. Snap and Pinterest sell against the same advertising budgets, so anyone covering one covers the other. Both add stock-based compensation back when they show adjusted profitability, and both describe it in almost identical language. Here is what it takes out.

EXHIBIT 2: The same adjustment, two companies, two years

Company	Fiscal year	Revenue (\$Mn)	Stock compensation(\$Mn)	Share of revenue
Snap	2023	4,606	1,324	28.7%
Pinterest	2023	3,055	648	21.2%
Snap	2024	5,361	1,041	19.4%
Pinterest	2024	3,646	766	21.0%

Figures as tagged in each company's Form 10-K and retrieved from the SEC's public company-facts service. Percentages calculated from those figures.

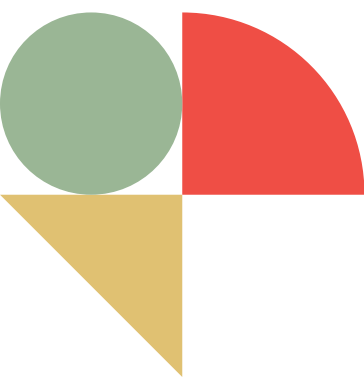
In 2023 that identical adjustment removed close to 29 cents of every revenue dollar at Snap and 21 cents at Pinterest. A year later the order had flipped: Snap's charge fell to 19.4% while Pinterest's drifted up to 21.0%. Neither company changed its definition. The size of the thing being removed simply moved.

So, a system comparing adjusted margins here is not comparing how well each business ran. It is comparing that, plus a different slice of removed cost for each, and the slice changes every year. An analyst carries this in her head; nothing in the data carries it anywhere. And this is the easy case, because both companies made the same choice. Where one excludes the charge and its competitor does not, the gap is not the difference between the amounts. It is the whole amount.

None of this is anecdotal. Research summarized by Deloitte finds that companies define earnings, interest, depreciation, and amortization differently, so the results do not compare, and that standard definitions do not exist. EY's review of SEC staff comment letters for the year ended June 30, 2025, put non-GAAP measures second among the most commented-on topics, behind only the narrative discussion companies write about their own results, with the share of registrants receiving those comments up more than 10% year on year.

AI makes this worse. Give a model an ambiguous measure and it does not stop to ask. It takes the most likely reading, which in financial text is usually the naive one, and writes it up in fluent professional English. Not obviously wrong. Confidently wrong, and it looks like good work. One analyst misreading a definition produces one bad comparison. A pipeline produces every affected comparison at once, with a clean-looking audit trail underneath. And a retrieval system built without a definitions layer does not fail loudly. It produces plausible output nobody will sign, surfacing 18 months later as a program that cannot ship.

A correct number delivered without its meaning is not a small defect in the product. In the research business it is the whole product, failing quietly.



What is it worth? Use your own numbers

We are not going to quote an industry-wide figure, because those rarely survive contact with a skeptical CFO. Do the arithmetic on your own desk. Take analyst headcount, multiply by fully loaded cost, then by the share of their time spent working out what reported figures mean before any view is formed. Thirty analysts at \$250,000 amounts to \$7.5 million a year; at 15% that is a little over \$1.1 million on work that produces no view and nothing a client sees. Run it again at 10%, or 5%. If the number stops justifying a quarter of anyone’s attention, that is a real answer — and a cheap one.

Analyst time is also the least interesting of the effects: review effort, onboarding, reconciliation cycles, duplicated work across research, risk and finance, and the share of AI drafts published without rework all move too. Baseline whatever you measure before you start.

Why not just buy this from a data provider?

It is the obvious shortcut, and for many purposes it works, which is why those products sell. But the provider owns that restatement, and two providers will not agree. More to the point, whether a charge is genuinely recurring is a house view — a research judgment about how a company makes its money, and part of what your clients pay for. A definitions record does not compete with vendor data. It sits above it, noting which basis each figure arrived on and where your house view differs.

A Record of What Your Measures Mean

3. The master record you do not have

No investment firm would operate without a security master. When several desks refer to the same bond, somebody keeps the master record: what it is, what terms define it, how it links to issuer, currency, and venue, and what changed when. Without one, two desks eventually trade what each believes is the same instrument and find out at settlement. The same argument applies to measures, and almost nobody accepts it. When analysts, models, and client feeds all refer to adjusted earnings or leverage, there is usually no master record of what those include, exclude, or can fairly be compared against. The result is the quiet cousin of a settlement break — two analysts comparing what each is sure is the same margin.

EXHIBIT 3: The record every investment firm keeps, and the one it does not

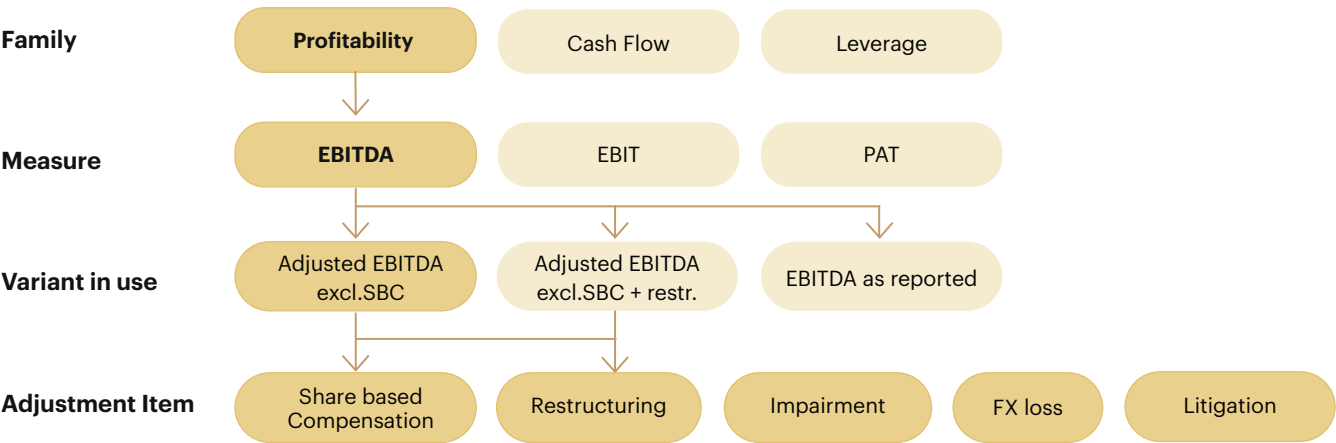
	Security master	Definitions master
Unit of record	One entry per instrument	One entry per measure
What it holds	What the bond is, its terms, its links to issuer, currency, venue	What the measure counts, what it excludes, which comparisons hold
Who owns it	Operations, with a named data owner	Nobody, in most firms. It belongs in research
What breaks without it	Two desks trade what each thinks is one instrument	Two analysts compare what each believes is one margin
How the break surfaces	Loudly, at settlement, the same day	Quietly, inside a published comparison, months later
Status in a typical firm	Universal, unarguable, funded	Absent, and not yet recognized as missing

The analogy holds closely, which is why this is not a novel proposal. The difference is historical rather than logical. Instrument ambiguity caused breaks that were loud and same-day; measurement ambiguity causes slow failures people absorb.

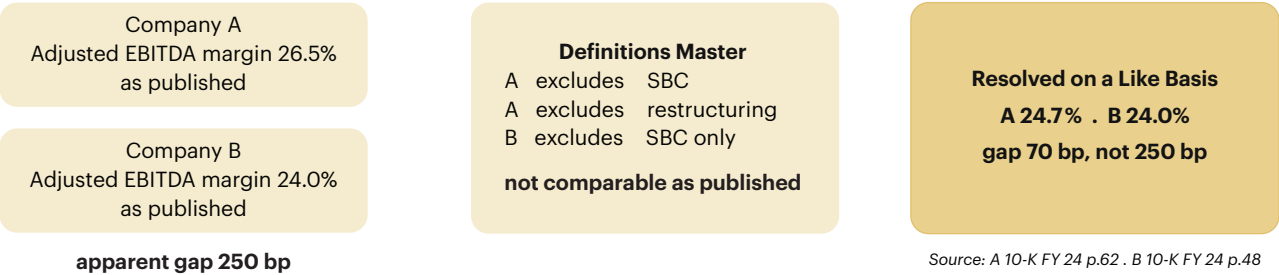
So what does that record look like? A **knowledge graph** holds three things: what a firm cares about, the named links between them, and the rules those links imply. Here the things are measures, companies, and the items that modify measures. The links are statements, such as this measure leaves out that item, or this figure came from that page. The rule falls out of them: if one company's figure excludes an item that another includes, the two cannot be compared until somebody closes the difference.

EXHIBIT 4: Inside the record: The structure, and one comparison resolved

The Taxonomy . an endless list becomes a finite hierarchy



The Graph at Work . one comparison, resolved



The System does not return a number. It returns a number that can be defended - with its basis, its adjustment and its source attached.

The taxonomy is what makes the problem finite. You cannot govern an open-ended list of ambiguous terms, but you can govern a hierarchy of measure families, the measures inside each, and a library of adjustment items that create the variants, because a new variant then slots into something that already exists.

And the output changes shape: Without the graph, the system hands you a number and you supply the meaning; with it, the number arrives carrying its basis, its adjustments, and its source.

One clarification. A graph stores the rule; it does not enforce it. **Enforcement** comes from a checking layer on top, which is well-trodden engineering and cheap next to everything else. What is not cheap is keeping the record true. That checking step is also what makes the system useful in a way most AI deployments are not. Before a figure enters a comparison, both sides get tested against the record, and what comes back may be a refusal: these two are not comparable as published; on a like basis the gap is 70 basis points rather than 250, and here is the adjustment. In research that may be the most valuable sentence a system can produce, because it is the moment a wrong conclusion gets stopped instead of published.

On cost, plainly

The software is mostly open source, which tempts people into calling this cheap. Licenses are the small part; the real cost is two people, one owning the definitions and one approving changes, for as long as you publish. Fund the build but not the upkeep and it decays within a year.

4. Could you do this with what you already own?

Some of it, probably. Several of these could hold definitions if you built enough on top of them, so the question is not whether it is possible. It is how much you would have to build, and how long before you have a usable answer.

EXHIBIT 5: Eight tools, one question: Where does meaning actually get settled?

Tool	What it is genuinely for	Meaning	What you would still have to add
Vector search	Finding text that reads like your query	○	Definitions, rules, sourcing, and ownership — all of it
Retrieval-augmented generation(RAG)	Grounding a model in your own documents	○	It finds the sentence that states a definition; it cannot apply one
Agent framework	Running multi-step work reliably	○	It runs an invalid comparison as obediently as a valid one
Enterprise search or data fabric	Finding and moving data across the estate	○	Both move figures; neither says what a figure means
Catalog or business glossary	Publishing definitions people can find and read	◐	Structure, versions, and rules a system can act on rather than retrieve
Semantic layer	Standardizing your own metrics for reporting	◐	Room for a company's own basis alongside yours, plus the comparability rules
Relational database or MDM	Storing governed reference data reliably	◐	Traversal, versioning, sourcing on each statement, and the comparison rules
Knowledge graph	Holding things, the links between them, and the rules those links imply	●	A checking layer to enforce the rules, and a named owner to keep it current

○ not what it is built for · ◐ part of the way · ● built for exactly this.

The first four are the AI stack most firms bought first; the last four are what the data function owns. Read the last column, not the third.

The dot in the bottom row is what a graph is designed for, not what you get on day one: an empty graph, or one nobody keeps current, is worth nothing, and it is the shortest route rather than the only one.

"We already have retrieval" is not an answer to this problem. It answers a different problem, finding the right document, which was solved first because it was easier.

“A good glossary would do most of this.” For a firm with one publishing format and a stable coverage list, quite possibly. But most glossaries hold narrative guidance written for a person to apply. A pipeline can retrieve that text, since modern catalogs expose glossary terms through APIs, but it cannot apply it. The distinction is narrative against structured, versioned, and resolvable.

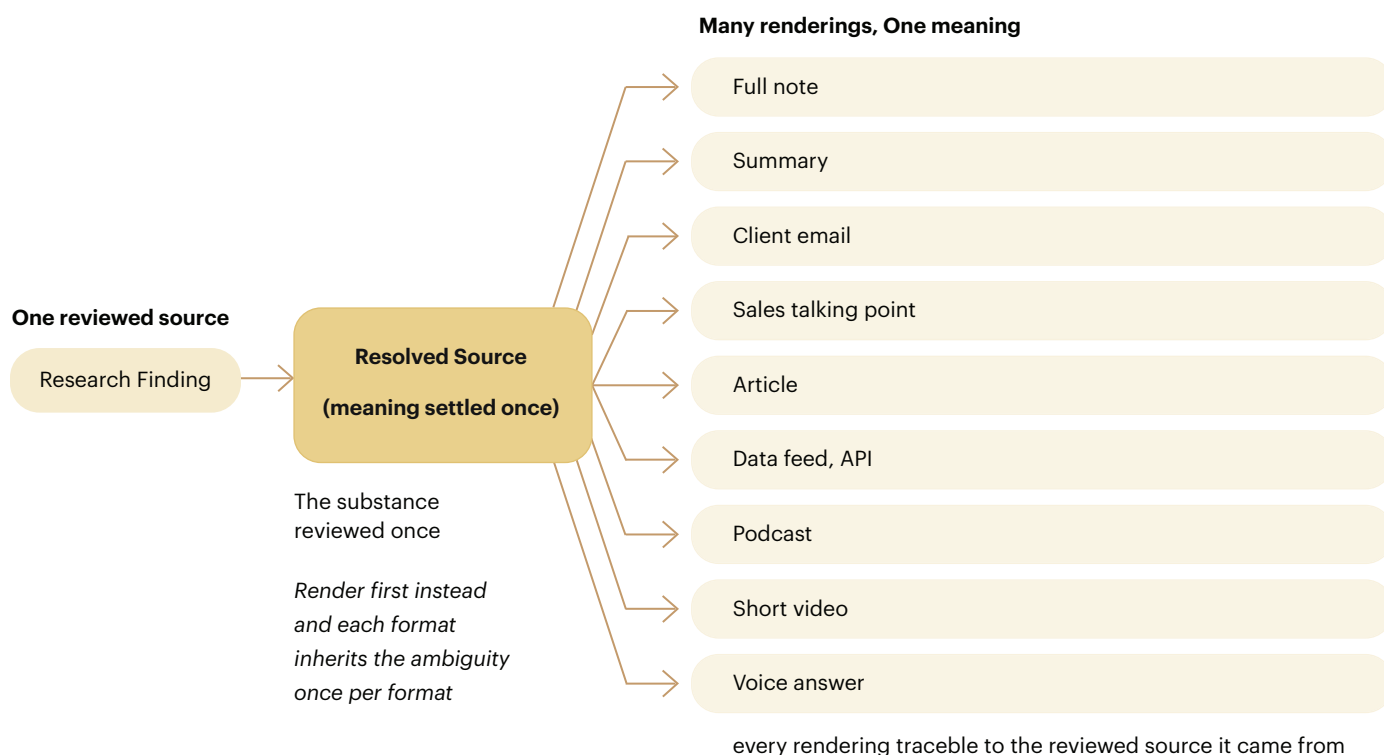
“We tried this before and it failed.” Also fair. Master-data programs and industry ontologies produced 20 years of elegant artifacts and thin adoption. But look at how they failed: filling the model in and keeping it current was manual, expensive, and never finished. The thinking was sound; the economics were not. Neither objection says the idea is wrong. Both are about price, which one measure in a quarter will settle.

Get Generation Right and Publication Follows

5. One source, many formats

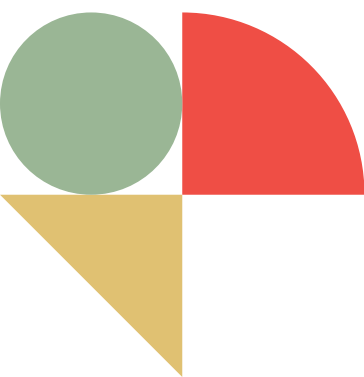
The same finding now reaches clients as a note, a summary, a client email, a talking point, an article, a data feed, an API response, a podcast, a video, and a spoken answer, most produced by hand today and inconsistently. Once the report itself is unambiguous, that becomes a rendering problem rather than a meaning problem, which is what a content factory does well.

EXHIBIT 6: One resolved source, rendered many ways



The order of operations is the whole argument. Resolve the meaning first and every format inherits it. Render first and each format inherits the ambiguity, once per format.

Audio and video make the point sharpest. A note can park a qualification in a footnote and a feed can carry it in a field, but nobody can speak a footnote. If the meaning is not settled in the source, those formats simply drop it, which makes them the hardest test of whether it ever was. Each derived piece is still its own communication, and each channel keeps its own compliance check, but the substance is settled once and every rendering traces back to it.



Getting There

6. What changed, what to do, and when to stop

What changed is the cost of filling the model in. A language model can read a filing and propose the entry: this measure excludes that item, here is the page that says so. A person checking proposals costs a fraction of one writing them, and that is why this is affordable now and was not before. What nobody can tell you yet is how accurate that extraction will be on your filings and your measures. How much review each entry needs decides whether this scales, so measure it in the first attempt rather than assuming it.

Governance decides this, not technology

The graph is not hard to build. It is hard to keep true, and that is an ownership problem. Whether the house view should exclude a charge is an analytical judgment about how a company earns its money, so it belongs to research rather than technology. Ask a technology function to settle it and they will either settle it badly or decline, leaving the definition unresolved while production routes around it.

Three mechanisms, none complicated: **a named owner** for each contested measure, senior enough that a ruling holds; **change control**, so a shift in convention is dated and applied everywhere; and **dating**, so last year's research can still be reproduced this year. The cost is a fraction of one role. Funding a build gets you a demonstration; funding stewardship gets you a capability, and a firm that will not name owners should not start.

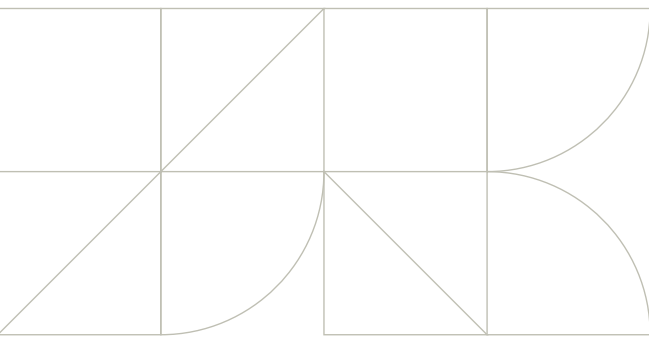
What a CIO is actually deciding

If nothing is done, no failure is visible. There are AI pilots that just demonstrate well and never ship, because no analyst will sign figures they cannot defend line by line. The business case was never accuracy. It is whether the work can be defended before a client, a reviewer or a regulator, which decides whether a capability reaches production. You cannot audit a prompt. You can audit a path through a definitions record, which also serves risk, finance, and client reporting.

A minimum viable pilot: One measure people genuinely argue about. Five to ten covered names, one quarter, two or three engineers, and a modest share of one senior analyst's time. Fundable from a research budget without a capital case. No new platform, no migration, no firm-wide model. Small is the point: it gives a real answer rather than a promising direction.

What success looks like: The measure resolves across every covered name, each adjustment traceable to source. The system declines a comparison it should decline and explains why. A named owner settles a real disagreement. And a figure published at the start of the quarter reproduces at the end, on the definition in force then.

When to stop: This matters more than the success criteria, because pilots rarely die of failure; they die of extension. Stop if no analyst prefers the resolved figure to their own spreadsheet, if no owner can be named, or if definitional disputes prove rarer than expected. And stop if 90 days becomes six months, because a slipping schedule is the first signal the scope was wrong. A firm that runs this and stops has learned something specific about its estate in a quarter, far cheaper than learning it three years in.



How Zensar Helps

Two capabilities, matching the two halves of this paper. Both are built to sit on top of what a firm already runs. The warehouse stays. The vendor feeds stay. The publishing platform stays, and so do the models your analysts have refined over years.

ZenseAI Context Engineering

This is where the knowledge graph models live. We build and maintain the record of what your measures mean: the measure families, the measures inside them, the variants actually in use across your coverage, and the library of adjustment items that create those variants. Agents built on that record resolve the definition before they use a figure, so what comes out carries its basis and its source. It works alongside the catalog and semantic layer you already own instead of asking you to replace them, and the record itself is held in an open standard.

Content Factory

The same finding has to reach clients as a note, a summary, an email, an article, a data feed, an API response, a podcast, a video and a spoken answer. Content Factory renders all of them from one resolved source, each version traceable back to the one that was reviewed. Audio makes the point sharpest: a note can park a qualification in a footnote, but nobody can speak a footnote, so if the meaning is not settled in the source, those formats simply drop it. The order matters more than either capability. A content factory built on ambiguous source material industrializes the ambiguity. Built on a resolved

EXHIBIT 7: What 90 days looks like

	WEEKS 1–2	WEEKS 3–8	WEEKS 9–12
What we do	Pick the contested measure with your analysts, baseline what it costs you today, agree on the success tests	Build the record for five to ten names, machine-proposed and human-approved, measuring the review each entry needs	Run live comparisons, show the refusals, reproduce a Q1 figure on the Q1 basis
What you provide	One senior analyst part-time, and a named owner for the measure	Access to filings and to your publishing stack	A decision: continue, or stop
What you end up with	A baseline you can measure against later	A working definitions record, and a real number for the review effort	Evidence in a room, not a status report

What you keep, and what we cannot promise

The definitions record is yours, in an open standard, and it runs without us. There is no platform to license and nothing to migrate off if you decide to stop at the end of the quarter. And we cannot tell you today how accurate machine extraction will be on your filings and your measures, because nobody can. Measuring that honestly is one of the things the 90 days is for.

What this is really about

The basis of competition in research keeps moving. It was access, then speed, then judgment applied on top of commoditized data. The next one is plainer: whether a firm’s output means what it says clearly enough that it cannot be misread. Every research organization will make this decision over the next couple of years, whether or not anyone frames it as one, because the systems commissioned this year will either settle meaning before they reason or they will not.

Model capability is no longer what holds this back. What is missing is a record of what the numbers mean, and at most firms that has never been anybody’s job. Ninety days and one contested measure will tell you whether yours can do this work.

Glossary

Semantic debt: The gap between what a firm's numbers say and what they mean. People pay the interest by hand for years before anyone touches the principal.

Non-GAAP measure: A measure the accounting standards do not define, such as adjusted earnings or free cash flow. How a company presents one is regulated; how it calculates one is not.

Definitions master: One authoritative entry per measure, as a security master holds one entry per instrument.

Knowledge graph: A record of things, the named links between them, and the rules those links imply, running from measure family to measure to the variants in use and the adjustment items behind them.

House view: The firm's own judgment on a contested question, such as whether a charge is genuinely recurring. Research work, not data work.

Content factory: Turns one settled source into every delivery format, instead of building each by hand.

References

Deloitte, On the Radar: Non-GAAP Financial Measures.

<https://dart.deloitte.com/USDART/home/publications/deloitte/on-the-radar/non-gaap-financial-measures>

EY, SEC Reporting Update: trends in SEC staff comment letters, year ended June 30, 2025.

<https://www.ey.com/content/dam/ey-unified-site/ey-com/en-us/technical/accountinglink/documents/ey-secru28204-251us-09-11-2025.pdf>

Snap Inc., Form 10-K FY2023 and FY2024.

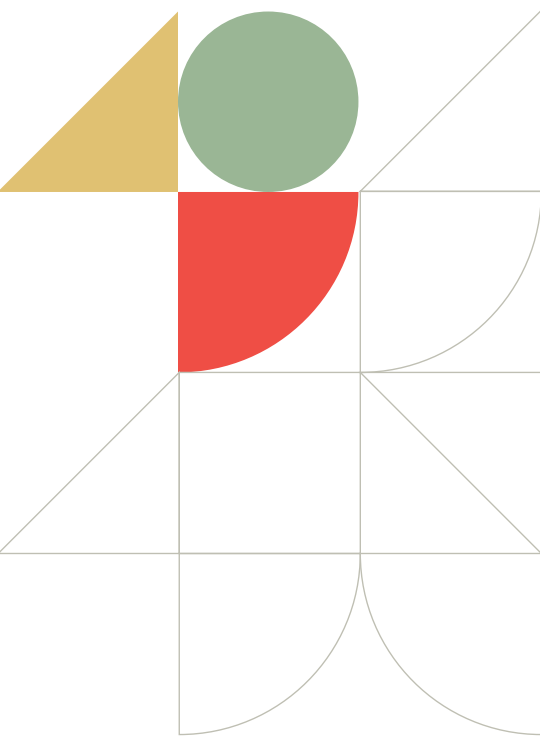
<https://www.sec.gov/Archives/edgar/data/1564408/000156440825000019/snap-20241231.htm>

Coinbase Global, Inc., Q4 2024 Shareholder Letter (Adjusted EBITDA definition revised in Q1 2024; prior periods recast).

<https://www.sec.gov/Archives/edgar/data/1679788/000167978825000021/q424shareholderletter.htm>

Pinterest, Inc., Form 10-K FY2023 and FY2024.

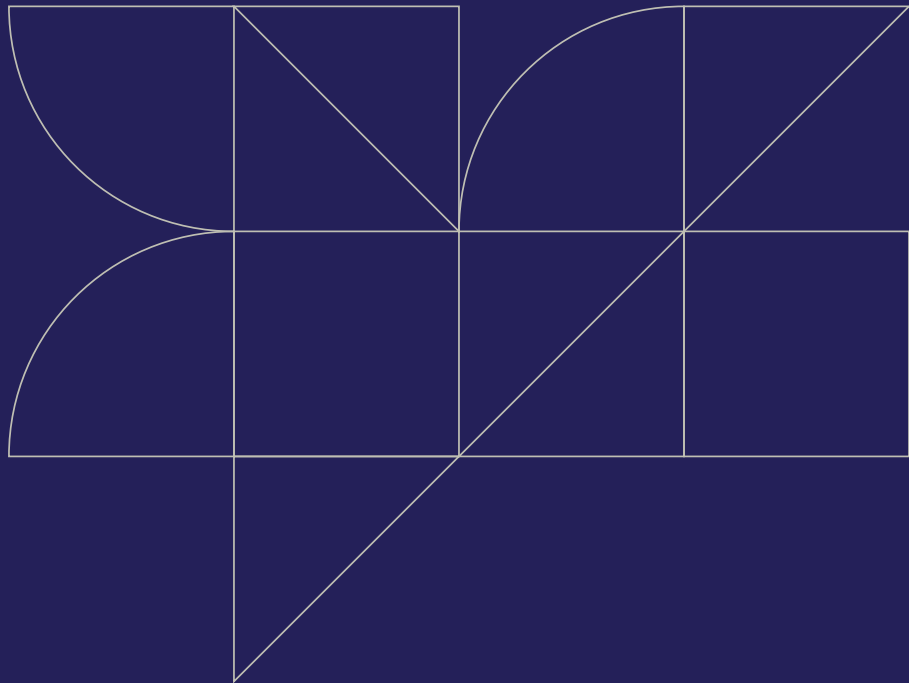
<https://www.sec.gov/Archives/edgar/data/1506293/000150629325000022/pins-20241231.htm>



Author

Kamlesh Kosare

Global Head of Business Consulting,
Banking & Financial Services - Capital Markets,
Industry Solutions Group (ISG)



zensar
An  **RPG** Company

At Zensar, we're 'experience-led everything.' We are committed to conceptualizing, designing, engineering, marketing, and managing digital solutions and experiences for over 145+ leading enterprises. Using our 3Es of experience, engineering, and engagement, we harness the power of technology, creativity, and insight to deliver impact.

Part of the \$4.8 billion RPG Group, we are headquartered in Pune, India. Our 10,000+ employees work across 30+ locations worldwide, including Milpitas, Seattle, Princeton, Cape Town, London, Zurich, Singapore, and Mexico City.

For more information, please contact: info@zensar.com | www.zensar.com