

DAKOTA DATA SUMMIT

---

# The State of Data & Distribution in Asset Management

Five patterns from a closed-door gathering of senior data and distribution operators at Dakota's Data Summit, Philadelphia.

The firms winning the next five years of distribution are not the ones with the best AI. They are the ones fixing the plumbing underneath it.

---

PUBLISHED

April 2026

FORMAT

Synthesis Report

SOURCE

[dakota.com](https://dakota.com)

## INSIDE THIS REPORT

# A synthesis of the day, organized around five patterns

---

No single firm has the complete playbook. That is the most important finding of the day, and the most freeing one for anyone reading this. The distribution function of the future is being built in real time, and nobody is as far ahead as the outside looks suggest.

What follows is Dakota's synthesis of the day, organized around the five patterns that emerged most clearly across the four panels, and what they mean for any firm trying to build a durable data and distribution capability.

---

**01**      **Building the Data Infrastructure Behind Distribution**  
Why the LLM is replaceable and the data is not, and what "AI-ready" actually requires.

---

**02**      **Sales Enablement at Scale**  
What it takes to get segmentation working, drive adoption, and measure pipeline impact.

---

**03**      **The Private Markets Distribution Machine**  
Where the CRM breaks down, and why the structural failures matter more than the technical ones.

---

**04**      **Where Data & Distribution Are Headed**  
Why governance is the real accelerant, and why compensation is the lever firms are pulling now.

---

**05**      **The Dakota View: You Are Not Buying a Tool. You Are Training a Mind.**  
A reframe for any firm rethinking its data and distribution stack in the agentic AI era.

---

## PANEL 01

# Building the Data Infrastructure Behind Distribution

---

## LLMs have changed user expectations, and not always for the better

The rise of LLMs has trained users to expect an answer, not a reference. Whether the answer is right is a separate problem, and one most distribution organizations are not yet equipped to handle. The dominant approach across the room: train models to return information from defined sources, and explicitly not answer when the underlying data is insufficient. Letting a model fill gaps produces confident wrong answers, which is worse than no answer at all.

The harder problem underneath is an education gap. Teams don't know what data their tools actually have access to, so they can't calibrate trust. The most disciplined firms are publishing internal changelogs of what their AI tools can and cannot answer, so users know what to rely on and what is still coming.

Worth naming: data no longer means relational tables. Pictures, voice, files, emails, all of it is now data. That shifts how infrastructure has to be built, and who owns it.

## Results engineering: define the outcome first

The panel converged on one principle: start with the business question, not the tool. The most common failure mode is dashboards with every metric possible and zero adoption, because nobody defined what success looked like. The fix is twofold. First, equip people with the right inputs, files, CRM, unstructured data, sales leads, and let them become forward engineers. Second, match what you're asking someone to do to the time they actually have. Empathy for the user is part of the architecture.

## What "AI-ready" actually means

The most concrete framework of the day. AI-readiness requires three specific things:

- Cloud storage with auto-metadata tagging on your files.
- A YAML file for every relational dataset that defines basic semantics: what AUM means, who the client is, how many definitions of a term exist across the org and which one wins.
- A skills layer that adds context specific to your business, so the model understands "we", your organization, your departments, your internal language, not just general knowledge.

The deeper point: get the lowest data level right first. Most firm logic exists as branches; the work is bringing it into a trunk so it's all in one place and you can grow consistent output from consistent data. Standardizing definitions matters more than picking the right tool. The number of tools available doesn't matter if your data isn't semantically clean underneath.

## Managing technical debt when the stack changes quarterly

Most leading shops are using both ChatGPT and Claude, not because one is better, but because they have different purposes. Cost is trivial; asymmetric upside justifies the redundancy. The principle: get the data right and you can swap the model. Clean data, clean connectors, and an agentic layer get you 80% of the way there with any major commercial model. The LLM is not the durable investment. The data is.

The hedge against tool flux is data governance. There is no way to predict what tool you'll be using in three months. Bringing it back to your data foundations is what makes that easier. And on decision-making itself: you can't tell the future, so you might be wrong, but you shouldn't be afraid to make the decision. Whatever makes that process faster is better. It takes bravery.

### PANEL 02

## Sales Enablement at Scale

---

### Segmentation: the long road to usable

Every firm on this panel described a years-long journey to get segmentation into a workable state. The starting point was the same in each case: the data foundation had to be rebuilt before segmentation was even possible. For some firms that meant blowing up the CRM entirely and starting over. For others, it meant building a top-down model with an outside partner, refusing to inherit someone else's definition of opportunity. In one case, the new model surfaced an entirely new category of unsegmented market the firm hadn't been addressing at all.

The common pattern: segmentation that runs across all channels (RIA, PWM, IBD), feeds marketing targeting and territory design, and is continuously refined by sales-team field insights. Top firms are tracking penetration rates, next best action, and specific behavioral signals, including whether reps are driving traffic to the firm's own website, which correlates positively with results.

## Driving adoption: what actually works

- **Executive sponsorship first.** Without top-down mandate, adoption fragments. Monthly scorecards create consistent accountability between managers and reps.
- **Explain the model's logic.** Wholesalers won't trust what they don't understand. Transparency about how segmentation scores are built is a prerequisite for use.
- **Work territory by territory.** Time spent with individual reps explaining the data, giving them the ability to adjust segmentation with a documented reason, and highlighting specific cases where data-driven targeting generated real opportunities. Showing the work is the best adoption tool.
- **Fill gaps with third-party data.** When internal segmentation is weak in a particular channel, supplement with external data rather than waiting to perfect it internally. Several firms in the room flagged this as the single fastest accelerant.

## What to actually measure

- **Pipeline quality and velocity** emerged as the consensus metric. Less lagging than AUM, more actionable than activity counts. Some firms also track redemption velocity as a leading engagement-risk indicator.
- **Sales lift**, tracking segmented advisors against unsegmented as a direct test of whether the model is generating results.
- **Don't let perfect be the enemy of good.** You don't need data perfect before finding successful ways to use AI.

## PANEL 03

# The Private Markets Distribution Machine

---

## Where private-markets data is structurally different

The most provocative position of the day: the long-term goal should be to eliminate the CRM entirely. The argument is that looking at data through a distribution lens biases it structurally. Letting distribution shape your vision of the data is the mistake. Be agnostic. One firm offered a concrete win from this approach: email bounce rate went from 13% to under 0.5% over three years as data hygiene improved.

Other firms in the room are taking different approaches to the same problem. Some operate across multiple CRM instances globally, with a strategy to find the single best instance of each data point and prioritize breadth. Redemption data is increasingly being positioned as an engagement trigger for the sales team. Family offices remain a grey area across the industry, with no clean categorization yet.

For firms operating in private wealth, the data is inherently messier than institutional. On the institutional side, it's clear who works with whom. On the wealth side, the complexity multiplies. The firms making the most progress here are doing so by mapping their manual processes in detail before automating any of them.

## Where technology fails

The panel was direct: the failures aren't primarily technological. They're structural.

- **CRMs built for traditional sales cycles** are rigid around vintage structures, earnings calls, and the unstructured data that now dominates private-markets workflows. Many firms work around them externally rather than inside the system.
- **Pipeline management works on the institutional side.** It breaks down on the retail/wealth side, where data complexity is higher and marketing-material transfers are harder.
- **CRM volume is a core limitation** for private wealth at most firms. Many are using dashboards or external tools to handle what their CRM can't.
- **Reconciling external sources to CRM records** remains the most manual part of the workflow. The goal everywhere is automation, but getting there requires first documenting exactly what the manual process is.

## PANEL 04

# Where Data & Distribution Are Headed

---

## AI is accelerating workflows, not transforming strategy

The clearest reframe of the day: AI is accelerating existing workflows, not transforming them. Strategic decisions about where to hire, which product sets to expand, and how to allocate distribution resources are still human decisions. AI is not doing that yet.

The honest version is that AI is the latest wave of FOMO. It's most valuable precisely where it meets an existing, defined workflow, meeting prep, post-meeting follow-up, email drafting, curating insights at scale. But the data underneath those workflows has to be clean. One firm shared an internal audit finding that just 1.7% of meeting data was being captured in their CRM. The industry average referenced in the room: 20 to 30%. Even that is a weak foundation for AI-assisted distribution.

For firms focused on ETF distribution, AI is being used to anticipate what an advisor or allocator wants and to drive segmentation decisions. The 13F data layer is central to that work on the institutional side. The growing challenge: third-party data sets are constantly changing frequencies and monetization models.

## Governance is the real accelerant

Governance is easy to skip over, and for some teams it's viewed as unnecessary. The firms dedicating real time to it are benefitting greatly. Without data governance, AI amplifies noise instead of signal. Firms that have invested in proprietary data quality, especially sales data, have a durable advantage that third-party data alone can't replicate. The framing in the room: you can get all the third-party data you want, but without the proprietary sales piece, you're not really winning.

The other half of governance is process. The distribution process itself needs to be mapped more granularly before AI can help at scale. Define who does what, what tools each step requires, what standard of data entry is expected, then hold people to it. AI works on a clean process; it amplifies a broken one.

## The compensation moment

One stat that stopped the room: meeting data entered the day after a conversation loses 46% of its character count. Two weeks out: 80% is gone. The response across leading firms is a decisive shift from inspire to require, building CRM compliance directly into discretionary bonus structures. The cultural conversation is hard. The logic is simple: it is paramount to explain to people what the future looks like; the technology we can't even envision is not that far away. Firms still operating on a voluntary-entry model should expect to be structurally behind within 12 months.

### THE DAKOTA VIEW

# You Are Not Buying a Tool. You Are Training a Mind.

*A reframe for any firm rethinking its data and distribution stack in the agentic AI era.*

---

Across the four panels, one pattern surfaced more often than any other: the model is replaceable, the data is not. Firms running ChatGPT and Claude side by side. Firms designing data flows that survive a CRM swap. Firms naming proprietary data quality as the only durable moat. The conclusion is the same in every framing.

That conclusion has a sharper edge than most firms have grasped. Not long ago, choosing an LP, GP, or private company database was a tool decision. You evaluated coverage, checked the price, and moved on. You could try it for a year and switch if it didn't work. The stakes were bounded. That decision is now one of the most consequential strategic choices a firm can make, because in the agentic AI era, you are not choosing a database. You are choosing what intelligence your entire organization will be trained on.

*"A database you chose wrong, you replace at renewal. A model you trained wrong, you carry forward, because it has learned, deeply and durably, from whatever you gave it."*

**GUI COSTIN · FOUNDER & CEO, DAKOTA**

## From procurement decision to training decision

The old frame was procurement. Which database gives us the best coverage at the best price? Evaluated annually. Switched when a better deal came along. Stakes bounded by the renewal cycle.

The new frame is training. What data are we training our model on, and what will it never be able to see? The decision compounds daily. It shapes every answer your AI will ever give. A wrong pick is not a slow quarter for analysts; it is twelve months of partial training that cannot be unwound by switching providers next year.

This is why the database decision has become the training decision. You are not choosing what data to have available in a search interface. You are choosing the cognitive raw material from which your AI's entire understanding of the private markets universe will be built. That understanding compounds, for better or worse, with every passing quarter.

## The only question that now matters

Once you accept that this is a training decision, one question becomes the most important one in the room: what percentage of the global LP, GP, and private company universe is your model actually being trained on? The answer determines what opportunities your model can surface, what prospects it can identify, what fundraising signals it can detect, and the ceiling of how smart it can ever get.

# 20%

### U.S. RIAs & Family Offices Only

Single channel, one geography. Common starting point.

# 40%

### U.S. & U.K. Institutional LPs

Two geographies, institutional focus. Misses most of the global universe.

# 80%

### Multi-Region LP Database, No GPs

Broad LP coverage, no GP, fund performance, or private company data.

# 100%

### Dakota, Full Global Universe

258,955 LP accounts. 26,080 GPs. 642,000+ private companies. Daily refresh.

The gap between 20% and 100% is not a coverage gap. It is an intelligence gap. A model trained on 20% of the market does not give you 20% of the answers. It gives you complete answers about a 20% slice of the world, with total silence on everything else. Your team does not know what the model doesn't know. The opportunities that live in the 80% it cannot see pass by invisibly, surfaced instead by a competitor whose model was trained on the complete picture.

## What this means for the firm reading this report

The conclusions from four panels translate into six things any firm can act on now:

- 1 You are not behind.**

The most sophisticated firms in the industry described themselves as somewhere between the middle and end of their own journey. No one has finished.
- 2 Fund governance before you fund tools.**

A dedicated senior data-governance role is the highest-leverage investment available. Firms that have made it are accelerating; firms that have not are stuck in tool evaluations.
- 3 Architect for unstructured data.**

Distribution intelligence will increasingly come from meeting notes, recordings, emails, and documents, not rows and columns. Build connectors and pipelines accordingly.
- 4 Stop treating the CRM as fixed.**

Design the next 18 months of decisions assuming the CRM's role will be renegotiated. Build data flows that survive a CRM swap.
- 5 Tie compensation to data hygiene this cycle.**

Every firm moving fastest has already done this. A voluntary CRM regime produces voluntary results.
- 6 Choose your training data with the seriousness of a strategic decision.**

The few thousand dollars that separate a partial database from a complete one is not the decision. The decision is whether the intelligence system your team will rely on every day, that compounds its knowledge daily, is built on a fraction of the market or all of it.

### ONE PREDICTION

In three years, the distribution organizations that matter will look less like sales teams with tools bolted on, and more like **integrated data operations with sales talent inside them**. The firms that make that transition early will spend the next decade compounding the advantage. The firms that wait will spend it catching up, one missed opportunity at a time, learning what it costs to train on the wrong data.

Dakota powers the foundation of that transition for asset managers across the U.S., Europe, and the Middle East. **Global LP, GP, and private company intelligence**, refreshed daily, connected to your AI through Claude, ChatGPT, or your own stack via Dakota's MCP server and Data API. To see how Dakota fits, reach out at [dakota.com](https://dakota.com).