INTELLIGENCE SERIES · 13F DATA

The 10 Biggest Problems With 13F Filing Data — And Why Almost No One Has Solved Them

The SEC's 13F database is one of the most valuable — and most brutally difficult — datasets in institutional finance. Here's why accessing it accurately is far harder than it looks.

Every quarter, thousands of institutional investment managers file a 13F with the SEC— disclosing their long positions in ETFs, closed-end funds, BDCs, and listed equities. On the surface, it sounds simple. The data is public. The SEC publishes it. Just go get it.

The reality is something else entirely. The SEC's EDGAR system houses one of the largest, most complex, and most poorly structured raw data archives in existence. Ingesting it, cleaning it, normalizing it, updating it, and making it actually useful
is a gargantuan undertaking that has defeated many who have tried.

At Dakota, we spent years conquering this problem — so our clients don't have to think about it for a single second. Here are the ten biggest reasons why 13F data is so extraordinarily difficult to work with.

THE 10 PROBLEMS

01 The Sheer Volume Is Staggering

There are more than 5,000 institutional filers submitting 13Fs every single quarter. Each filing can contain hundreds or thousands of individual line items. Multiply that across 20+ years of history, and you are looking at hundreds of millions of data points that must be ingested, parsed, and stored with perfect fidelity — every quarter, like clockwork. The SEC's EDGAR system does not serve this data in a clean, queryable format. It serves raw XML and text files that
must be parsed at scale. Volume alone disqualifies most organizations before they even get to the hard problems.

02 Filer Identity Is a Complete Mess

The SEC assigns a CIK (Central Index Key) to each filer — but the relationship between CIK numbers and actual investment firms is far from clean. The same firm may file under multiple CIK numbers due to mergers, rebranding, holding company structures, or administrative errors. Subsidiaries file separately from parent companies. A single RIA with multiple affiliated entities may generate dozens of separate filings that need to be stitched together into one coherent picture of
the firm's total positions. Without deep entity resolution work, your data looks
like thousands of anonymous fragments rather than a map of institutional ownership.

03 Security Identification Is Inconsistent and Unreliable

13F filers are required to use CUSIP numbers to identify securities — but in practice, the data is riddled with errors. CUSIPs are mistyped, outdated, or simply wrong. The same ETF may appear under multiple CUSIPs across different filers. Share classes are frequently confused. When a fund changes its structure, ticker, or CUSIP, filers don't always update consistently, creating phantom positions and broken time series that corrupt any analysis built on top of the
raw data. You cannot simply trust the CUSIP. Every security identifier must be validated against a live reference database and corrected where necessary.

04   The 45-Day Lag Creates a Stale Data Problem

Filers have 45 days after the close of each calendar quarter to submit their 13F. That means the most recent data you can ever see reflects holdings that are already six to seven weeks old by the time the filing is made — and many filers submit on the last possible day. Tracking and flagging late filers, amended filings, and withdrawn filings requires continuous monitoring of EDGAR for changes, not a simple quarterly pull. If you miss an amendment, you are reporting positions that the filer has already corrected.

05   Amended Filings Silently Overwrite History

When a filer submits an amended 13F — a 13F/A — that amendment supersedes the original filing. But the SEC does not send notifications or alerts. If your data
pipeline is not continuously monitoring EDGAR for amendments, you will retain the original (incorrect) data while the filer has already corrected the record. In a
dataset used to track institutional positioning, stale or corrected data that
hasn't been updated can completely distort the picture of who owns what — and by how much. Amendments happen far more often than most people realize, and missing them is the difference between intelligence and misinformation.

"Bad time series data doesn't just produce wrong answers. It produces confident-looking wrong answers, which is worse."

06   Confidential Treatment Requests Create Invisible Holdings

The SEC allows filers to request confidential treatment for certain positions —typically when disclosure would reveal a material non-public investment strategy. These positions are simply absent from the public filing. There is no placeholder, no indication that data is missing. For anyone trying to build a complete picture of a firm's holdings, these invisible gaps are nearly impossible to detect unless you have a methodology for identifying them. A filer's total reported market value may be dramatically understated if significant positions are under confidential treatment.

07   The Raw Data Has No Context — Just Numbers

A raw 13F tells you a firm name, a CUSIP, a share count, and a market value. That's it. It tells you nothing about the firm's investment mandate, its AUM, its channel, its decision-makers, its other holdings, or its investment philosophy. Without a rich contextual layer wrapped around the filing data, you cannot answer any of the questions that actually matter: Is this a meaningful position for this firm? Is this firm a target for our distribution team? Who do we call? The raw data, in isolation, is nearly useless for anyone trying to make decisions with
it. This is precisely where Dakota's proprietary RIA firm and contact database transforms 13F data from raw numbers into actionable intelligence.

08   Normalizing Position Sizes Across Vastly Different Firms Is Non-Trivial

A $10 million position in an ETF means something entirely different for a $200 million RIA than it does for a $20 billion pension fund. The raw 13F data gives you no denominator. Without knowing the filer's total AUM — which is not in the 13F — you cannot contextualize any position. You cannot identify overweights, underweights, or meaningful concentrations. You cannot rank holders by conviction. Every analysis that requires normalized position sizing requires an external source of AUM data that must be mapped back to each filer with precision.

09   Building Reliable Quarter-Over-Quarter Time Series Is Extremely Hard

Tracking how a firm's position in a given fund has changed over time sounds straightforward — but it requires flawlessly linking the same filer, the same security, and the same position across every quarterly filing, accounting for
entity name changes, CIK migrations, CUSIP changes, fund reorganizations, and amended filings. A single broken link in the chain corrupts the entire time
series. For any product that claims to show institutional flow data — whether a holder is adding, trimming, or exiting — the quality of that time series is everything, and building it right is enormously complex. Bad time series data doesn't just produce wrong answers. It produces confident-looking wrong answers, which is worse.

10   Keeping It All Current Requires Relentless, Ongoing Infrastructure

Many organizations can get a 13F dataset to a usable state once. Almost none can keep it continuously updated, validated, amended, and enriched at production quality over years and years. The SEC's filing calendar doesn't stop. Filers change. Firms merge and rebrand. New filers enter and old ones exit. Funds reorganize. CUSIPs expire. Every one of these changes must be captured, resolved, and propagated through the entire dataset in real time. The infrastructure
required to do this — and to do it right — is not a one-time project. It is a permanent, full-time commitment. At Dakota, this infrastructure has been built, battle-tested, and is running continuously — so that every query our clients run reflects the most current, most accurate picture of 13F data in the market.

THE BOTTOM LINE

Dakota Conquered Every One of These Problems. So You Never Have To.

Dakota Marketplace combines fully normalized, continuously updated 13F filing data with the most comprehensive RIA firm and contact intelligence database ever built — and layers AI on top, so you can ask any question and get an answer in seconds.

See how Dakota Marketplace handles 13F data.

Request a demo and see the data for yourself.

REQUEST A DEMO