




# Hive

## The Hive Mind Guide To: Historical Data

In partnership with:



**BMLL**



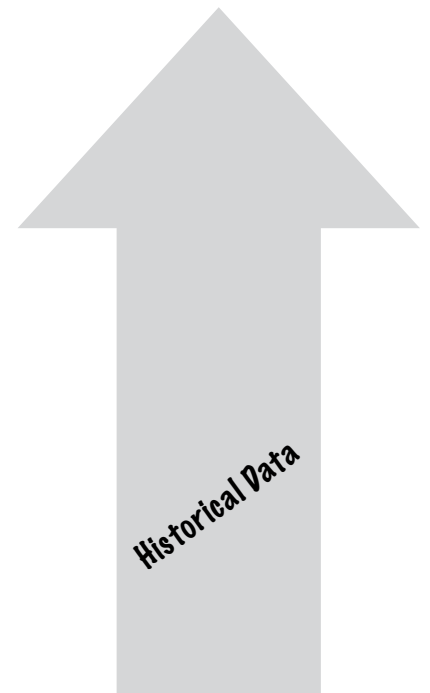
Bringing together cutting-edge insights from The Hive Community, complemented by benchmarking data, enriched with real-world experience, and offering practical guidance to help you navigate this rapidly evolving landscape.

# The Problem

In today's trading landscape, nearly everyone has access to real-time data. As a result, firms must now seek new avenues for gaining a competitive advantage. The challenge lies not in accessing immediate market updates but in leveraging **quality historical data** - clean, well-formatted, and reliable - to uncover deeper insights that drive smarter decision-making.

Historical data forms the bedrock of informed trading strategies, offering critical insights into market trends, price movements, and execution quality. However, despite the widespread availability of real-time data, many firms struggle to integrate historical data effectively at the point of execution. The underlying issues of rising costs, fragmented sources, and inconsistent data quality persist.

In an environment where every player can access up-to-the-minute information, the true differentiator is the ability to harness historical data to forecast trends, optimise algorithms, and refine execution strategies.

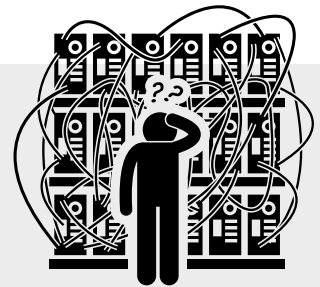


# Current trends

**Automation Driven by Data Quality:** As AI and machine learning models become central to trading, historical data is more critical than ever. Even the best algorithms will fail if the input data is flawed.



**Integrated Data Platforms:** The evolution toward a “single source of truth” within EMS/OMS environments underscores the need for consolidated, clean data streams. Firms that integrate their data directly into these systems experience significant reductions in latency and errors.



**Visualisation and Analytics:** Tools such as Tableau and PowerBI are increasingly used to create intuitive dashboards. These allow cross-functional teams—from trading desks to compliance—to access and act on data insights rapidly.



**The Promise and Pitfalls:** Initiatives like the consolidated tape aim to unify fragmented data sources. However, concerns persist about the accuracy of data at the source and whether increased costs will translate into real liquidity benefits.



**A Call for a “Golden Unicorn”:** Market participants have long sought a platform that seamlessly aggregates and cleanses data across asset classes. Although systems are moving in this direction, the challenge remains substantial.



# Pitfalls

## Cost vs. Value and Economies of Scale

- **Rising Costs, Uneven Access:** While data costs continue to climb, the value derived from it isn't always clear. Many smaller firms find comprehensive access cost-prohibitive, whereas larger institutions can leverage economies of scale.
- **Budgetary Cycles:** Firms often experience boom-and-bust cycles in data spending. During lean phases, even critical data—like that needed for pre- and post-trade analyses—may be compromised due to cost constraints.
- **Two-Tier Pricing and Provider Dynamics:** With various data vendors (from established giants to more niche providers), firms frequently face unbundled or re-bundled pricing structures that complicate cost assessments. The challenge is compounded when trying to pass these costs onto clients.



## The Imperative of Data Quality

- **Foundation for AI & Automation:** As trading becomes increasingly automated, the reliability of AI models depends entirely on the quality of historical data. Incorrect or interpolated numbers (a common issue even with leading providers) can undermine strategy performance.
- **Cross-Asset Consistency:** Whether it's equities, fixed income, distressed credit, or commodities, data quality remains a linchpin. In asset classes where historical records are incomplete or estimates are the norm, extra diligence is needed.
- **Real-Life Impacts:** Proprietary, anonymised insights reveal that incorrect data can lead to mis-pricing and missed opportunities. Even advanced analytics like TCA only deliver value when built on a solid data foundation.





# Historical Data in Trading: From Insight to Execution

## Bridging the Pre- and Post-Trade Divide

### Pre-Trade Estimates

Advanced trading platforms are increasingly seeking to offer pre-trade cost estimates for every line in the EMS. This enables traders to anticipate execution costs and refine strategies on the fly.

### Post-Trade Analysis

Detailed post-trade reviews, such as those seen in TCA platforms, help in comparing the estimated costs against actual execution. Over time, this aids in calibrating models and validating strategies.



### The Real Cost of Data

As one industry voice noted, *“Cost is the number one concern.”* Without clear, consolidated historical data, efforts to predict trading outcomes can falter—highlighting the need for a unified view that spans all asset classes.

## Proprietary Insights and Real-Life Use Cases

- **Internal vs. External Data:** Many firms now operate with a hybrid model—maintaining an internal data repository enriched with external feeds from trusted providers. This dual approach enables them to leverage proprietary insights while maintaining breadth and depth in data coverage.
- **Dynamic Data Integration:** Some platforms are being used not just as execution systems but also as integrated data hubs. In practice, these systems support everything from generating new data columns on the fly to running granular, asset-specific TCA reviews.

**Market-Specific Challenges:** For example, in the distressed credit and convertible bonds markets, the lack of reliable historical data has been a significant barrier. Hive Members in global equities, ETFs, futures and options report spending hours manually cleaning and reconciling data—an effort that underlines the importance of having high-quality inputs from the start.

# Common Data Challenges (and Finding the Right Solutions)



## **Data That's Messy, Incomplete, or Outdated**

Clean, complete data is the foundation for accurate insights. Prioritise providers who deliver fully normalised, cross-venue data on a T+1 basis to avoid time-consuming data prep.



## **Unclear Methodologies Undermining Confidence**

Insights are only as good as their data source. Push for providers who offer clear, standardised methodologies so you can trust the analytics driving your decisions.



## **Data Management That Eats Up Time and Resources**

Managing multiple datasets in-house is costly. Offload data procurement, licensing, and management to a trusted third party to focus your resources on generating insights, not handling data.



## **Fragmented Data Mapping Across Asset Classes**

Inconsistent identifiers slow down analysis. Choose platforms offering seamless mapping across ISIN, FIGI, and other identifiers to create clear, cross-asset insights faster.




## **Rigid Workflows Limiting Analysis Speed**

If your platform can't plug into your preferred tools (like Python, R, or SQL), you'll hit efficiency roadblocks. Opt for solutions offering flexible access through APIs or cloud platforms.



## **Limited Internal Data Science Capacity**

Without strong in-house data science resources, extracting insights can be slow. Partner with providers offering hands-on support to help accelerate your analytics capabilities.



**Pro Insight from BMLL:**  
Accessing high-quality, normalised historical data on a T+1 basis helps firms benchmark broker performance and optimise execution strategies — without the data clean-up burden.





# Taking the Next Steps

## Assess Your Data Infrastructure

Review how historical data is collected, stored, and integrated into your trading systems. Identify gaps and areas for improvement.



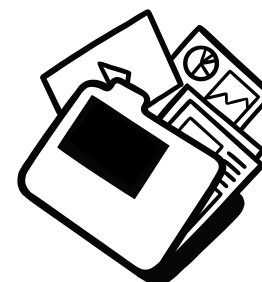
## Invest in Quality

Prioritise providers that deliver clean, well-formatted data. Remember, the payoff from advanced analytics and AI begins with high-quality inputs.



## Integrate and Centralise

Aim to consolidate data streams across asset classes within your EMS/OMS. A unified data platform not only improves decision-making but also helps in scaling advanced trading strategies.



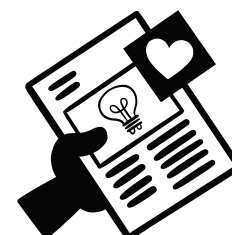
## Validate Across Conditions

Test trading models in varied market regimes. Use rolling windows for backtesting to ensure that strategies remain robust in both bull and bear markets.



## Foster Collaboration

Encourage ongoing dialogue between trading, data analytics, and compliance teams. This synergy is crucial to unlocking the true value of historical data.





## About The Finance Hive

The Retail Hive brings buy side heads of trading together to soundboard, ask questions, benchmark and collaborate as a collective. They do this by providing safe, co-owned and curated environments for trading heads and partners to build meaningful relationships, give one another value, identify solutions, and promote best practice.

[www.thehive-network.com/the-finance-hive](http://www.thehive-network.com/the-finance-hive)



## About BMLL

BMLL Technologies is the leading, independent provider of harmonised, Level 3, 2 and 1 historical data and analytics to the world's most sophisticated capital market participants, covering global equities and ETFs, as well as global futures and US equities options.

BMLL offers banks, brokers, asset managers, hedge funds, global exchange groups, academic institutions, and regulators immediate and flexible access to the most granular Level 3, 2 and 1 T+1 order book data and advanced analytics. BMLL gives users the ability to understand market behaviour, accelerate research, optimise trading strategies, and generate alpha more predictably.

[www.bmlitech.com](http://www.bmlitech.com)

