

Financial data management – A start to a new approach

Monali Samaddar and Ankur Jain

Ensuring a good data management process is a priority for today's financial services organisations, since businesses across the globe are increasingly relying on their data to power their everyday operations. The level of data sophistication has risen significantly in the past 20 years. As the frequency of user interactions and the volume of data stored increase and change rapidly, there is a need to identify the challenges in [data management](#). But the buck does not stop here. It is imperative that organisations build an adaptable, yet robust, framework to manage these challenges. The framework could be built internally or in collaboration with experienced firms that create bespoke solutions

The usual suspects:

Growing data volumes

Large volumes of data are created every day, and organisations face the continued challenge of aggregating, managing and creating value from all this data.

Poor data quality

Data stored in structured databases or repositories is often incomplete, inconsistent or out of date. Missing data items may be due to errors in downloading or the company being analysed not having reported the data item. Often, data from multiple data vendors may show differences in historical time-series data and/or lags in data availability, making it more difficult to derive insights. As data vendors evolve and transform their business, they tend to change their formats or stop publishing periodically due to maintenance or other reasons. Many format changes are announced with very short notice, while others are not announced at all. Additionally, vendors may revise the data from time to time if the initial value was not correct (point-in-time data refers to the numbers reported by the company initially, whereas subsequent changes in financial data are referred to as restated data). For any analysis, we need to ensure we have all the available information.

Data integration

With varied structure of data, dispersed over different channels, probably the biggest challenge comes in the form of integrating the collected data. At present, most of the data acquired is unstructured and of no use until or unless it is structured. For example, data is most often available in horizontal format with no proper data formats, making the process of data analysis more time

consuming. In addition, the data available with vendors is generally in different base units (e.g., thousands, millions, or billions), requiring manual checking and rebasing the data uniformly for comparison.

Reporting-frequency mismatch

Some companies, for example, those in the UK, Australia and the Netherlands, may publish their financial statements on a semi-annual basis, as they are not required to publish quarterly. In the US, however, the SEC requires all companies to publish quarterly. This mismatch in reporting frequency may make analysis of global data difficult.

Using improper primary keys

The primary key should be selected with care. Analysts mostly use a stock's ticker as the primary key, which may be incorrect (especially when used without the country suffix), as there could be multiple companies in the different exchanges with the same ticker. Furthermore, in certain scenarios, such as a company's name change due to a corporate action, the ticker may change, making it difficult to perform a historical time-series analysis.

Use of grouping keys

When conducting macro-level research, we frequently use stock-level data and aggregate the metrics based on group identifiers. Without a primary grouping key, it would not be possible to aggregate the data into categories such as sector and industry.

Duplication of data

While calculating aggregates, care should be taken to avoid including duplicate data points. For example, we should consider removing tickers that have multiple classes of shares, as it would otherwise double count the same company's financial metrics. For example, Berkshire Hathaway (BRK) has two classes of shares (Class A and Class B) listed on the NYSE. When we download stock-level data for the S&P 500 from any data vendor, we will get values for both share classes. As the underlying company is the same, line items such as sales, net income and EBITDA would be the same for both classes of shares. While aggregating, we could mistakenly aggregate both classes, resulting in double counting the same metric.

The experienced cop:

To enable better decision making, business processes and operations need to analyse the potential problems and come up with solutions both at the organisational and individual level. We discuss below a few plausible solutions to the problems we have highlighted.

Handling the four Vs of data

To address the challenge of the four Vs of data - volume, velocity, veracity, and variety - firms could build large enterprise data warehouses. They could also look at automation and analytical tools to improve the data handling and storage processes.

Poor data quality

Inaccurate data is often the most difficult data-quality issue to spot. Ensuring clear procedures consistently can be a good practice. Using automation tools and data validation checks could also help to reduce the amount of manual work required to check for mistakes. While working with point-

in-time data and restated data, checks need to be run. For example, when conducting a returns analysis, we should consider data that was available as of or before the period of computing the returns (which explains the price movement), rather than using the latest restated values. Missing/incomplete datasets can also be tricky to deal with since it is not always possible to back-calculate and arrive at a number relating to a previous period. However, one way we can progress is by putting in threshold checks such that our outcome should be meaningful. For example, the CUSIP code has nine digits; therefore, if the field contains fewer/more than the required number of digits, the analysis would generate an error.

Data integration

To reduce complexity in data integration, companies need to combine their existing and new data into a single data universe. The solution to transforming unstructured into structured data could lie with automated systems that can actively classify the unstructured data, adding system metadata and user metadata, and eliminating any encountered redundancy to successfully transform it into structured data.

Reporting-frequency mismatch

To make numbers comparable in the event of reporting-frequency errors, we could use an approach such as linear interpolation to arrive at quarterly data using semi-annual data.

Using primary keys

Avoiding the use of improper primary keys could help save precious time that could instead be used to create meaningful insights. Where applicable, we could strive to use more robust primary keys such as the FactSet Identifier (FSID), MSCI Time-Series Identifier (TSID), SEDOL identifiers or CUSIP code. Such an identifier would be unique to a company, making it robust for conducting global analysis. It would also not change in the event of a company's name change due to a corporate action.

Using grouping keys

When using grouping keys, we should consider standardised grouping keys such as GICS sector classification to maintain consistency across countries. This would be useful in correctly aggregating data based on sector.

Duplication of data

To overcome this problem, we need to identify data points that would help us retrieve a unique data set, or we should build validation engines or checkpoints to remove duplication. For example, in the event a secondary class of share is removed, we need to pull another field that categorises different share classes, and based on our requirement, we could either keep or remove the secondary class of share or listing on multiple exchanges.

Conclusion

The data challenges discussed indicate that a robust framework is required to deal with them. However, the approach has to have elements of adaptability since we are dealing not only with multiple data points but also with different data challenges – from accuracy to format to sources.

At Acuity Knowledge Partners, we employ Data Bridge to tackle data challenges. This is a robust, flexible, bespoke data solution that we develop for our clients to automate most of their [investment](#)

[research](#) and reporting data workflow management. This bespoke data solution follows a technology-agnostic approach and broadly covers four aspects of the data workflow: (a) aggregation, (b) validation, (c) analytics and (d) reporting

Data Bridge is developed by our integrated team of experts in investment management, data and technology, which is one of the key differentiators. Improved data accuracy, productivity, reduced turnaround times and ability to scale are some of the common benefits that our [asset management](#) and hedge fund clients have gained by deploying Data Bridge. We believe an automated data solution such as Data Bridge will play a critical role in the financial management world, going forward.

About Acuity Knowledge Partners

Acuity Knowledge Partners is a leading provider of high-value research, analytics and business intelligence to the financial services sector. The company supports over 350+ financial institutions and consulting companies through a team of over 3,000+ subject matter experts who work as an extension of the clients' teams based out of various global delivery centres.

We empower our clients to drive revenues higher. We innovate using our proprietary technology and automation solutions. We enable our clients to transform their operating model and cost base.