Inside Market Data

November 2013

ANALYTICS

SPECIAL REPORT

Sponsored by:

DATAWATCH

PLATTS
McGRAW HILL FINANCIAL

THOMSON REUTERS

Visually Analyze Any Data, At the Speed of Business

In today's financial markets, trading firms must leverage every shred of data for competitive advantage. And with the amount of data continually growing, it's critical for firms to analyze, understand and interact with this data regardless of its variety, velocity or structure.

Datawatch is the leading provider of visual data discovery solutions that allow trading firms to optimize the use of any information whether it's structured, unstructured, or semi–structured data locked in content like static reports, PDF files, and EDI streams in real–time sources like CEP engines, tick or other historical sources. Through an unmatched visual data discovery environment, Datawatch allows firms to utilize ALL their data to deliver a complete model of the trading environment in real time.

Datawatch provides the solution to Get the Whole Story!

Learn more and download your FREE trial today at www.datawatch.com or contact sales@datawatch.com.

DATAWATCH | GET THE WHOLE STORY™

271 Mill Road, Quorum Office Park, Chelmsford, MA 01824, USA

Buy-Side Technology **Awards 2013**
waterstechnology
Best Big Data Technology Provider to the Buy-Side
Winner
Datawatch Corporation

Charting the Next Generation of Analytics

Understanding volumes of market data has always been a challenge for traders, asset managers and investors attempting to make sense of numbers flashing across a screen. Hence, over the years, some of the most effective data platforms have been those with analytical tools that help traders make sense of that data, that add context, and provide insight into more than just the in-the-moment price changes. After all, a single price out of context and with no reference to its movement has little meaning, and one of the best ways to display that movement over time is visually.

With the rise of algorithmic trading and the increase in machine-readable data inputs, some of these analytics became entirely computerized, relying on technologies that can crunch vast volumes of data to create their own context without the need of visual displays geared at human traders.

But as it becomes harder to compete in the high-frequency trading space, and as instruments and markets become more complicated, traders in markets with lower levels of automation have discovered new value in analytics' ability to provide insight where machines as yet cannot. And for many at the institutional level, the next stage of analytics will involve more use of black-box engines to perform complex calculations at low latency. But for many more—including those at both the institutional and retail level—while analytics will look very different from those of the past, they will continue to be visual processes: it's just that these will be more complex than before, to account for the fact that higher volumes of faster and more granular data exist today, and will serve the purpose of providing precision insight into these volumes of new data types.

And to accompany these new analytics, expect to see new types of delivery mechanisms to broaden their availability, such as the use of "app store"-like platforms operated by vendors that want to offer compelling analytics but are unable—or see no need—to build their own to compete with established providers. These platforms in turn will also provide a channel for specialist providers to gain a wider audience.

The next generation of analytics is full of opportunity. And taking advantage of the opportunity to leverage these new tools will help investors grasp opportunities in the marketplace that they wouldn't have spotted without them. ∎

Max Bowie
Editor, *Inside Market Data*

Inside Market Data

An incisive*media* publication

recycle
When you have finished with this magazine please recycle it.

Behavioral Indicator Vendor Titan TA Closes

Titan Trading Analytics, the Edmonton, Canada and Atlanta, GA-based provider of trading signals based on behavioral analytics, has shut down as a result of "financial difficulties" and an inability to meet its financial obligations.

In late August, Titan took down its website and replaced it with a code-verified portal leading to an announcement that the company had ceased all business operations, and that all directors and employees had either resigned or been made redundant.

Titan's TickAnalyst product was a hosted platform that ran a series of algorithms against 10 years' worth of tick data, volatility data and social media content to generate recommendations for individual stocks and industry sectors based on current investor "buzz" around an asset, and on historical price movements in response to similar circumstances in the past.

The closure comes despite the vendor signing two deals for TickAnalyst this year: one to Canadian investment advisor Jordan Advisory Services, and the other to Eddras Capital Management, a startup alternative asset manager in Atlanta, GA founded by three former Titan employees—Eric Davidson, Daniel Robinson and Andrew Saideman.

According to Titan's 2012 annual report, the vendor made a net loss of $2,269,669 for the year on revenues of only $35,157, and incurred expenses of $2,304,806, including almost $1 million in general and administrative costs, and more than $600,000 in research and development—despite a review of costs to reduce monthly spend to $50,000 by cutting salary and consultant fees, and despite its board of directors waiving any remuneration for the year.

In addition, the results note that the company no longer employed a direct sales staff or used distribution channels, instead relying on "established relationships" with money managers, and warned that it expected to "continue to incur losses and negative cash flow" and "may have to curtail or cease operations," if unable to commercialize its products. ■

Tradesignal Bows Version 7 of Data Platform

German data display and analytics vendor Tradesignal has released version 7.0 of its Enterprise Edition data and analytics workstation, which includes a new user interface to enable more efficient access to Tradesignal's functions.

Version 7 of Tradesignal also features a new mini-toolbar to allow users to make quick chart changes without searching through multiple menus, a new Debugger module to support in-house programming language Equilla as well as new chart tools, indicators and chart types. ■

eVestment Integrates Morningstar Data into Analytics Tools

Data from Chicago-based investment research and data provider Morningstar is now available via fund management software provider eVestment's analytics solutions suite, including its PerTrac and Fundspire systems.

The Morningstar data includes data on funds, annuities, insurance group separate accounts, Morningstar's category classifications, and the Morningstar MSCI Hedge Fund Indices, globally or by region, providing eVestment clients with more information to support modeling of the various investment vehicles available to investors. ■

S&P CapIQ Adds Credit Risk Indicators to XpressFeed

S&P Capital IQ is making its entire universe of pre-scored point-in-time, short- to mid-term and long-term credit risk indicators for financial institutions and non-financial corporates available via its XpressFeed data platform.

The suite of models, dubbed PD (probability of default) Model Market Signals, PD Model Fundamentals, and Credit-Model, provide indications of credit risk over a range of time horizons for almost 50,000 listed companies and more than three million private companies—730,000 of which have financial data available.

The PD Model Market Signals, which provide daily indications of credit risk based on factors such as credit default swap (CDS) spreads or stock prices that change on a daily basis, will go live on XpressFeed this week. Meanwhile, PD Fundamentals, which addresses short-to-mid-term credit risk by assessing the probability that an entity will not meet its obligations in the next one to two years, and CreditModel, which considers a three- to 10-year time horizon for creditworthiness, will both go live over the coming months. Each model can deliver credit risk scores based on the standard AAA ratings system or as a probability of default.

"It's very important for the models to be developed on the latest data, particularly as we have gone through a recent financial crisis. We are lucky in a statistical sense in that we have been in an unprecedented recession phase…. plus many industries have changed drastically over [the past] 10 years, so we don't think that data from beyond 2002 should be used in models that our clients use to predict the future," says Marcel Heinrichs, director of market development at S&P Capital IQ.

Previously available via a web-based interface, the credit risk indicators can now be delivered as a daily flat file via the XpressFeed datafeed service and integrated into firms' internal systems. ■

The Power of Visual Data Discovery

With traditional analytical and data visualization techniques no longer capable of handling and displaying the volumes of content associated with Big Data analysis, firms must invest in new infrastructures and visualization tools to achieve big benefits from Big Data, says Ben Plummer, chief marketing officer and senior vice president of strategic alliances at Datawatch.

The simplistic charting tools included in Microsoft Excel and most dashboard software systems are inadequate for presenting real-time data in ways that foster comprehensive understanding and support useful analysis in a short amount of time. The introduction of complex event processing (CEP) into a capital markets operation presents even greater challenges, since the quality of operational decisions made by traders and their managers have direct and immediate impacts on the bottom line.

People in capital markets firms must take many different factors into account, including risk, price changes, liquidity, and portfolio balance in order to make good business decisions. As data comes in from the markets and trading systems in real-time, traditional reporting and visualization tools are not up to the task of providing human decision-makers with the low-latency visibility they need to support timely and accurate decision making.

The introduction of CEP, in-memory columnar databases and related technologies into the infrastructure provides banks and trading firms with ever more intricate ways to implement complex, multi-dimensional business logic in the systems used to process tick data. In addition, these systems can be re-configured quickly to support new regulatory requirements or altered business practices. However, even though CEP engines can greatly reduce the amount of data that must be analyzed by human operators, their outputs are still real-time streaming feeds in most deployments. Users need tools that foster comprehensive understanding and support useful analysis in a short amount of time.

In some firms, managers still spend precious hours wading through paper reports showing yesterday's activity. Some firms have advanced beyond paper reports, but are still working with the equivalent of paper reports on their screens. Far too many financial service professionals spend valuable time—hours a day in many cases—trying to make sense of columnar data in Excel or as presented in on-screen reports from their trading platforms.

Another issue confronting managers is the tendency of users to create multiple unconnected data silos using tools like Excel. It is not uncommon to find a large number of spreadsheets on different machines or stored in different email messages that all contain similar but not identical information. This propagation problem can easily lead to costly misunderstandings and time-consuming confusion. A better approach utilizes a centralized, database-driven solution that supplies a single version of the truth. This solution also allows users to easily change the way the data is displayed, and allows users to interact with the data in ways that make sense based on the particular challenges and interests associated with their positions.

The most savvy banks and fund managers are now using real-time data visualizations in concert with their advanced data handling infrastructures. These visual data discovery solutions are optimized for real-time, real-world use cases, including risk management (investment, counterparty, credit and liquidity), transaction cost analysis, pre-trade basket analytics, and portfolio performance and attribution.

Big Data presents many technical challenges, as most tools on the market are unable to cope with Gartner's "Three Vs of Big Data" (volume, velocity and variety), as the magnitude of data in all three dimensions has increased—particularly since the recent financial crisis. Visual data discovery increases visibility to critical Big Data sources, offering broader analytical capabilities that allow people to cope with data sources of any size, that change in real time, and data stored in multiple types of systems and formats, whether structured, unstructured, or semi-structured.

The right tools will support several important activities, including:
- Real-time views of aggregated risk, tracking error, performance and associated attribution from all funds under management down to the constituent level of a single portfolio;
- Visualization of the impact of changes to underlying constituents;
- Calculation of returns for traded components on an intra-day basis;
- Customization for each portfolio manager's actual decision process;
- Measurement of the effects of decisions on portfolio attribution and risk;
- Collaboration within the organization.

Visual data discovery solutions provide traders and managers with clear and up-to-the-minute understandings of how investment choices and markets are performing. Further, this highly visual approach to data analysis allows users to understand precisely how a portfolio is performing and responding to changes in the market. Traders can identify problems and see the impact of changes, and can take immediate action to maximize profitability. Managers can see all details for all portfolios in a single view and drill all the way down to see details about the underlying instruments, or drill up to view aggregated data across different metrics. They can also analyze portfolio relative contributions, screen out constituents that do not merit attention, and bring up complete details on the ones they need to focus on. ■

Making Big Data Bite-Sized

The key to creating next-generation analytics is harnessing and managing the underlying next-generation datasets that make analytics essential and allow companies to create valuable tools to get the most value from data, says Asif Alam, global head of machine-readable news at Thomson Reuters.

The zettabyte era is approaching fast. According to the research consensus of numerous industry analyst firms, in the next year or so, the amount of IP-based data will exceed the storage capacity of 1 billion laptop computers, each with a 1-terabyte hard drive. The vast majority of this will be unstructured data.

As "big data" volumes continue to grow, people and computers have a harder time extracting useful insights from these larger and more diverse datasets. People buckle when trying to put the sheer amount of data into easily understood context. Applications, on the other hand, have trouble crunching unstructured data.

However, life improved for both when third-party automated sentiment analysis offerings started appearing on the scene.

By using these systems that convert the market's view on specific events or general topics into a positive, negative or neutral value, traders can incorporate market color and other variables into their automated trading strategies, which had previously been limited to structured market and historical data.

Meanwhile, analysts can feed the same sentiment data into visualization tools to create easily understandable and digestible charts and graphs, examining anything from political instability in a frontier market to detecting market bubbles in Scandinavian vacation housing, or even showing which celebrity is the flavor of the week among teenagers.

Risk managers can implement sentiment analysis to quantify event risk and manage portfolio volatility. Research analysts can use it to explain price movements of assets or specific stocks quickly. Trade support teams can incorporate sentiment data into stock screener applications for retail clients. Even compliance professionals can use it to identify market manipulation events.

The secret behind turning subjective unstructured data into objective, quantifiable sentiment data is computational linguistic analysis, which uses statistical or rules-based approaches to model natural language.

Until recently, the shortcomings of the three most popular analysis methodologies—lexical analysis, grammatical analysis and machine learning—made them ill-equipped to handle sentiment analysis due to their inability to handle ambiguity, slow processing speeds or complex implementation process.

Yet, a hybrid methodology has emerged that breaks down sentences into individual tokens that identify a word or phrase's grammatical function, comparing that to an extensive lexicon, before finally being classified using a large archive consisting of manually annotated samples. This method leverages the speed, stability and flexibility of all three methodologies without their shortcomings.

The Sky's the Limit

It's important to remember that sentiment analysis is far more than whether a piece of information is viewed as good, bad or neutral. It can also identify how relevant it is to the topic at hand and how it aligns with other similar information on the topic.

Once a piece of data has assigned values representing its polarity, magnitude and novelty, it can be viewed and manipulated like any three-dimensional object, like a brick. And you can make some pretty amazing things with bricks.

Earlier this year, Thomson Reuters launched the TRust Index that benchmarks sentiment for the top 50 global financial institutions based on market capitalization. The idea behind the index—which secures its data from an array of Thomson Reuters sources—is to use it as a proxy for trust in the financial services industry as a whole.

Imagination and creativity are the only limitations to organizations' ability to reap the benefits of tapping into the value that is locked in unstructured data sources for use by humans, not just machines.

Consider how much more accurate equities research analysts' forecasts can be if they implement similar models that quantify base emotions like fear, joy, optimism or violence occurring in the regions, industries and markets that they cover.

On a less macro level, a firm can create an index around a particular stock or security to include its market sentiment and those of its competitors. What might be good news for General Motors shareholders might not be good news for Ford or Toyota shareholders, for example.

Improving the stock screening process is just one of the lowest-hanging benefits of adopting such models. Improving counterparty risk surveillance is another. Instead of tracking a company and its rivals, an index could track the sentiment of a company and its relationship with its major and minor counterparties and business partners. Many in the industry probably wish they had this capability before Bear Stearns, Lehman Brothers, MF Global and Peregrine Financial Group all went down in flames.

Automated sentiment analysis might have started as an interesting novelty a few years ago, but since then has certainly become a necessity. ■

Trillium Begins Post-Trade Analytics Push

Trillium Labs, a software development spin-off from New York-based proprietary trading firm Trillium Trading is rolling out a new post-trade analysis tool for US equities trading that compares traders' order data to full order book market data to identify manipulative behavior, and plans to sell the tool to other trading firms to help them comply with anti-market manipulation rules.

Trillium already uses the tool, dubbed Surveyor, internally to monitor traders'' activity to identify "layering"—when a trader floods one side of the order book with orders that they have no intention of executing, which are intended to drive prices up or down, then at the same time places opposite orders on the other side

of the order book to execute against the artificially raised or lowered prices.

Each day, Surveyor creates a list of questionable trading from the prior day, ranked according to certain factors to denote the seriousness of the incident. Surveyor's screen is split into multiple panes. One shows the spread, along with details of orders, cancellations and executions. Another displays the total available volume in the consolidated US equities order book, with an overlay showing how much of that volume at any time was the activity of a single trader, and another chart that shows the volume of orders that trader is placing on the opposing side of the order book.

Compliance officers can use this data

to look at any trading deemed questionable, and either dismiss it or investigate it further. If, for example, a trader forgot to cancel orders on one side of the book while pursuing a separate strategy, the compliance officer may recommend that they adjust their trading practices to avoid being flagged in future. Or if the officer believes the trader was intentionally layering, he can discipline the trader and report the actions to regulators.

The firm built the system after being investigated for layering by the Financial Industry Regulatory Authority (Finra) between 2007 and 2010, which led to the firm being fined $2.26 million and firing nine traders, and being forced to set up controls to prevent future instances. ∎

State Street Adds Analytics Vets

State Street Global Exchange, the newly-formed data and analytics business of State Street, recently hired Jeffrey Bohn as head of portfolio analytics and valuation, and Thomas Hughes as senior product strategist for analytics.

Bohn was most recently managing director of PwC Japan, prior to which he was chief executive of Soliton Japan, head of portfolio analytics and economic capital at Standard Chartered, general manager of Shinsei Bank, and spent almost 10 years in senior roles at Moody's KMV. Hughes also spent over 14 years at Moody's as senior vice president, then served as president and chief operating officer at The Connors Group. ∎

Derivative Path Taps Fincad for OTC Analytics, Valuations

Derivative Path, a startup provider of automated trading technologies for over-the-counter derivatives, has incorporated Vancouver, Canada-based OTC derivatives pricing and risk management vendor Fincad's analytics platform into its outsourced technology platform for structuring and executing hedges, as well as for regulatory reporting, risk management and accounting support.

Officials say incorporating Fincad's analytics will allow Derivative Path to provide clients with advice and trade valuations based on accurate, real-time analytics, and to expand into new asset classes. ∎

Cowen Readies Predictive Analytics for Clients

New York-based investment bank Cowen Group is rolling out a suite of predictive analytics and strategy switching tools to support clients' algorithmic and electronic trading activities, to help the firm grow its presence in US institutional equities.

In the first phase of the initiative, Cowen rolled out a strategy switching engine and a short-term alpha model—which predicts short-term price movements—to its internal trading desk.

The switching engine contains basic logic to change strategies based on price movements and inputs such as regression. "We look at buy-side traders' activity via

their FIX messages, so we know how they are reacting in real time. The switching engine, BEST (Best Execution Switching Technology), will do what they're [already] doing," says John Cosenza, co-head of electronic trading at Cowen.

The model uses a range of data inputs, including large volumes of historical data, technical factors and correlations, and machine-learning techniques to derive a signal of where a price will be five minutes in the future, based on a rolling six-month dataset of as many data points as possible.

The second phase of development will productize these tools in a way that can be

made available to clients via a web portal, which will allow them to see Cowen's predictive analytics for orders they submit electronically. Cosenza says the firm will spend the next couple of months testing this capability internally before sending it into production later this year.

A third phase of the rollout—currently in the planning stage, with plans to enter production next year—would allow clients to access the analytics on-demand via the internet. Cosenza says the firm plans to allow clients to enter a security symbol into an online interface which would return a score for that specific symbol. ∎

Analysis Evolution Requires Analytics Revolution

For a long time, it was accepted that "data is data." But now, financial firms and investors are looking to incorporate completely new data types into pricing models, risk management and trading strategies. With this comes the need for new tools to analyze this data—both the new types available, and the sheer volume being generated—in new ways, which is generating a flurry of activity among traditional and next-generation analytics providers. The analytics of the future won't look like the old line graphs of the past. So what will they look like, and how will their value change as a result?

IMD: As trading evolves to become increasingly automated, how is the concept of "analytics" evolving from traditional charting displays to provide new types of indicators, signals, and calculations?

Asif Alam, global head of machine-readable news, Thomson Reuters: In the past, analytics consisted of real-time charting, market volume and stock quotes, allowing market participants to identify potentially costly outliers, and view historical intraday stock profiles. Those tools are still used today, but the analytics toolkit for a trader has expanded significantly with innovations geared towards providing personalized analytics tools that traders are looking for to set themselves apart from the competition.

Today, analytics are being used to find the signal in the noise across a variety of data sources from structured sources, like fundamental stock data, to unstructured sources, like news and social media. Innovations in the analytics space continue to re-define the boundaries of what can be obtained, how it can be consumed and even how it can be displayed for use by machines and humans alike.

Intelligent analytics tools, like sentiment analysis of news and social media, are increasingly becoming part of the workflow in a trading environment. These tools allow traders to set specific parameters of what sources they want to include (from price to news to social media), what to measure from those sources (from a sentiment measurement of trust or conflict to an increase or decrease in price), and finally, how they want the end result delivered (from numerical format to feed an algorithmic trading system to visualization for human consumption) to ultimately inform their trading strategy.

Catherine Turley, head of analytics, CIMB: Providing analytics for humans is fundamentally different to providing analytics for consumption by machines. Humans seek to evaluate numerical information in relation to other numbers, thus the interpretation of data is highly subjective, being dependent on the context that is presented with the analytic. For example, when studying graphs, the choice of appropriate timescales and granularity is important and can determine whether an outlier appears significant or not.

On the other hand, machines have no ability to interpret analytics abstractly, so instead of context, the metadata associated with the analytic becomes ever more important (in terms of confidence/ statistical distribution, etc.). This removes any subjective bias and should result in a more accurate understanding of the underlying data.

Because of the limitations of the average data consumer, traditional analytics were reduced to one-dimensional time-series. With the increase in computing power and technological sophistication, models can be written to accept multi-dimensional inputs and associated descriptive statistics. This ability to process multi-dimensional inputs has changed the way in which analytics are consumed.

Ben Plummer, chief marketing officer and senior vice president of strategic alliances, Datawatch: Traditional analytics for trading covered two areas. The first is the charting and graphics components bundled with—or added to—market data display terminals. The main problem with these is they lack the capability to actually analyze data. Rather, they give a graphical representation of the data in the moment, perhaps with a historical view.

The second area is portfolio and risk, where spreadsheet graphics have been used to track what happened on the desk and how that impacts the financial health of a firm. Unfortunately, this information is static and historical, so it can be hours or days after an opportunity or risk event occurs before it is understood—or worse, before a firm is aware that it occurred.

Next-generation analytics applications provide insight into all of a firm's data to help traders get at the specific opportunities that are available in the market as they occur. Likewise, portfolio and risk managers must have access to the same information as traders, at the same time, in order to understand the full context of a firm's opportunity and risk profile.

For analytics to provide value to a trading firm, it must provide true analysis of all of the firm's datasets, and it must be capable of analyzing that data in real time.

Renee Krall, commercial director, market data, Platts: Traditional charting packages and analytics very much look at price action on its own. However, the story in markets such as physical oil trading goes beyond price—for example, volume and market depth are equally as important. Getting the whole story, and then looking at behavioral and activity trends is where the market's analytics evolution is headed. For example, in the fuel oil markets, where you have the flexibility of submitting fixed or floating prices, understanding those signals and indicators, and looking at why and when those changes occur, can offer insights beyond those available by simply looking at price.

James Maclachlan, senior trader, CF Global: The use of analytics is quickly evolving to enhance alpha. Where a relevant data pool exists, we can observe with greater accuracy our choice of execution strategy. Market conditions, order size, timing and impact make up some of the components of TCA (transaction cost analysis).

Adam Honoré, CEO, MarketsTech LLC: This is a big question. For the most part, the capital markets industry is way, way behind in how we view and leverage data. Look at the recent news about Monsanto buying The Climate Corporation,

Ben Plummer
Chief Marketing Officer and Senior Vice President, Strategic Alliances
Datawatch
Tel: +1 978 441 2200
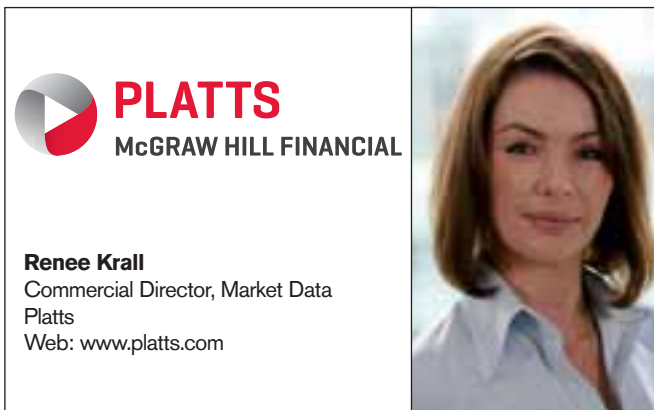Email: Ben_Plummer@datawatch.com
Web: www.datawatch.com

which correlates crop yield data, soil types, and weather data for every two square miles in the US. That cost nearly a billion dollars to buy. Someone made a joke once about something I wrote, where they suggested I was saying they should be able to compare elephant migration path data to commodity prices in Africa, and I'm thinking the fact that they think they're being funny about that means they just don't understand: that's exactly what they should be doing. An antelope herd took out 22 percent of my mother-in-law's sunflower crop last year, so who's laughing now? Old-school versus new-school is people with a myopic view of what constitutes a fundamental versus those willing to think data is only extraneous when it has been absolutely proven to have no value.

Paul Rowady, senior analyst, Tabb Group: We are definitely not in Kansas anymore. Though increasingly immersive visualizations of today—the "grandchildren" of the original charting solutions of the late 80s and early 90s—provide dramatically enhanced density of information flow to end-users, it is the coordination of other components of the technology stack that deserves equivalent praise. First, there is simply a whole new galaxy of raw data available. The capital markets ecosystem is entering into a phase of maturation where the idea of "global markets on demand" is becoming more real every day. Data from markets across the globe—developed, developing, emerging and frontier—and across multiple lit and dark liquidity pools can be accessed from a single API. But we certainly don't stop there. Full-spectrum asset class and product class coverage is included. On top of all that, the data is far, far more granular. Ticks and depth of market data is available in tiny sub-millisecond slivers of time for the most microstructure obsessive players.

And, of course, lest we bury the headline here, there are the words themselves. The massive and unprecedented sea of raw unstructured, semi-structured and polymorphic content—known by taxi drivers and grandmothers alike today by the unfortunate label, Big Data—that, by virtue of the internet and increasingly pervasive data standards like XML, is now sufficiently automated and digitized to be available for signal mining.

Automated access to all these numbers and all these words is what is driving the epic wave of interest in derived data, known as analytics. When you marry all of this new fodder with the

PLATTS
McGRAW HILL FINANCIAL

Renee Krall
Commercial Director, Market Data
Platts
Web: www.platts.com

equivalently impressive computational firepower available today to detect patterns, explore expansive combinations of the ingredients in a search for meaning, and otherwise boil the ocean in ways never before conceived, well, I think you understand now my reference to Dorothy and Kansas.

IMD: What does this reflect about how the roles of "consumers"—human or machine—at end-user firms are changing? What trends are driving these changes, and how are data providers innovating to address these requirements in detail?
Krall: In the past, our customers tended to be happy with us presenting them with the data. Now, there's much more demand from the customer to be able to really dig into every piece of information that went into assessing that final number. There is considerably more rigor demanded in how the data is collected and presented: expectations are far higher in terms of what they should have available to them, so that they can understand and trust what they are looking at. Our feedback has been that clients want to be able to access every piece of information that led to the final number, understanding the methodologies used in coming up with the assessment, and all the fundamental pieces involved.

Also, you cannot ignore the impact that the regulatory landscape is having on data, and on the requirements that regulatory imperatives impose on companies to capture, produce and present information on their market activity and behavior—both internally, up the chain, and externally.

We recognized some time ago that these regulatory requirements, and the way that the market needed to ingest and manage data, presented us with an opportunity to set the standard for commodity data and present our information in a completely consistent, accessible and centralized way. In the past, users might have been prepared to source data from several different places. Now they want to get all their data in one delivery mechanism, and one format.

Plummer: Prior to the financial crisis, trading firms' strategies were driven from the trading desk. It was very one-sided, and there was an assumption that the risk exposure from trading operations was negligible.

The financial crisis has driven firms to take a holistic view of the entire business as a component of the market ecosystem. Not only

does management need a view of the risk associated with a specific desk or trader in the context of their firm's overall risk, but it must also consider the risk factors of the firm in the context of what's happening in the market and economy.

Rowady: End-user skill profiles are changing, and need to change more. The super-driver here is the mass transition from manual tasks to automated tasks in global capital markets and financial services, given that the underlying data and computational intelligence is now becoming so much more fluent that the value added tasks for people are concentrating further downstream in the workflow (or higher in the "knowledge pyramid") as the machines replace them earlier in the workflow. Think of an automotive assembly line in the 1930s or even the 1970s compared to the assembly lines of today.

"Trigger pullers," "spreadsheet jockeys" and even "stockbrokers"—all labels from my early days in this business—have all but been relegated to the history books. What do these roles—of a trader, an analyst, or a broker—have in common? Their value proposition, in large part, was in aggregating and finding meaning in disparate data. Now that disparate data is more automated and normalized, and now that broad libraries of techniques for pattern recognition have been codified, and racks of silicon-laced engines stand ready to crunch gushers of bits, those former roles and skills are giving way to new roles and new skills.

Of course, there are still lifetimes worth of disparate data to be assembled and converted into intelligence that will hopefully keep folks like me employed for a little while longer. So, if human skills, in general, need to move higher up in the knowledge pyramid (as in, data-to-information-to-knowledge-to-wisdom—aka DIKW—hierarchy) then I expect that human-based roles in our business should migrate more towards business analysis, strategy, and design, in addition to the wave of quantitative engineers that we have already seen—in other words, any area requiring levels of creativity that are greater than current mechanical capabilities. (The design theme intrigues me the most, by the way. Someday, I envision a run on Xbox gamers on Wall Street, just as there has been a run on quants, of late.)

As for data providers, you should notice that more data is now more easily accessible through a single, unified application program interface (API). I expect this trend to continue, as these providers simultaneously aggregate more data and configure APIs to be sensitive to metadata—the atomic structure of the feeds themselves. The API is what allows solution designers to plug the firehose into the latest computational engines and exploratory graphical user interfaces (GUIs).

Turley: Data volumes have increased exponentially in recent years—an increase generally attributed to automated trading engines. Firstly, we observe an increase in the frequency of trade execution and order placement across stocks. Secondly, because of the ability of multi-dimensional models to process several input streams, new data sources and datafeeds are combined with traditional market data, to aid in alpha-seeking strategies.

Asif Alam
Global Head of Machine-Readable News
Thomson Reuters
Tel: +1 877 365 1455
Email: asif.alam@thomsonreuters.com
Web: thomsonreuters.com/site/trust/

While there has been innovation from data providers in the technology and speed at which this data is delivered to downstream applications, there has been less focus on the interpretation and normalization of this data across different exchanges, especially with regards to non-numerical data streams.

Honoré: The data providers are always innovating because most of them have their eye on another market. Most of it is social. The capital markets actually need to be more engaged in those efforts. For instance, we've all seen funds aggregate Twitter sentiment, and they're obviously still doing it because companies like Gnip and DataSift seem to be doing OK. But how much more valuable would Twitter be if someone in capital markets had suggested Twitter ask a few basic questions about each member so you could correlate very specific groups of people, like all IBM employees or all people in a specific profession? Right now, we're kind of like remora fish just swimming around attached to the sharks in San Jose, Boulder, Austin, etc. The other thing I'd say is that we have to change our view of trust. You would think we'd have learned from the ratings agencies in 2007 and 2008 that authoritative people can be very, very wrong. I talked to Estimize the other day, and their chief executive told me his biggest challenge is getting investment managers to accept that his data is valid, even though a portion of the aggregate may be coming from someone working in their basement. The world is changing, but we still consider data having more intrinsic value if it came from someone wearing a suit.

Maclachlan: The shift to analytics reflects a new sophistication in trading analysis demands. Alpha protection is key. Every basis point matters in in the new world of anemic returns. Unbundling has also sharpened the focus on execution. In addition, the buy side has been educated with a myriad of new tools created by electronic providers keen to differentiate themselves.

Alam: Consumers of analytics today continue to include both machines and humans, with a growing trend towards the development of more innovative tools geared towards the human consumption and use of analytics. Financial markets participants are looking for extremely robust analytics tools that give them the power to create their own analytics from a wide variety of datasets. With such proprietary analytics systems, unique real-time actionable insight in the form of a chart or graph allows human traders to easily find the signal in the noise, informing their day-to-day trading decisions and their broader trading strategies.

IMD: With such large volumes of market data, combined with emerging use of Big Data to derive value from many different datasets—from news, social media to consumer trends—is the underlying data itself still as important as the analytics, or have analytics themselves become the new source of "data" and signals?

Rowady: Yes and no. For machines, raw data will continue to be the primary fuel, by default. As raw data volumes have grown, they have become too large and unwieldy for productive human consumption, digestion and output. So, the machines will remain tethered to raw data and then additional layers of derived data, or analytics. Humans are becoming untethered from raw data so that the lowest level of interaction will become analytics.

But there is much more to the story here: While machines and humans will overlap in their consumption of derived data, humans will increasingly consume data through pictures (which ties in with my earlier point about design). At some point—and maybe even already—the sheer volume of analytics will become too vast for regular human consumption. More and more, the "output" will be in pictures first, and then only upon a need to verify the details, will there be drilling down into the layers of that data structure.

Turley: Arguably, analytics have always been more important than the underlying data, as they are of a higher order. This is ever more apparent as the breadth of datasets grows to include non-numerical and loosely structured sources.

Because of a machine's inability to abstractly interpret information, a news feed is unusable until normalized. The role of the analytics engine should be to understand and normalize data, whether this is a sentiment engine interpreting news into a numerical feed

Catherine Turley
CIMB

to be plugged into a model, or conflation of large volumes of granular order book data into summary analytics which maintain the structure of the underlying data.

This reinforces the requirement to design systems such that the appropriate technologies are responsible for appropriate specialization. Specifically, the role of market data vendor should be to provide the most granular data as fast as possible, normalized and cleansed. The role of the analytics engine should be to understand this data, and summarize it for input into trading engines, which should then be able to take advantage of new data streams without incurring prohibitive processing costs.

ROUNDTABLE

Plummer: Data is always trying to tell us something. When there was less of it, it was easier to get at what it was trying to say. The markets continue to generate increasing amounts of data. At the same time, there are many more data sources that are relevant to making decisions about trading and risk. All of this data is still trying to tell us something. However, with so much data, that message can easily get lost, or can come through too late. Next-generation analytics tools provide the ability to collate all of the data so that we're able see the message concisely as it is being communicated. These tools aren't replacing the underlying data, they're helping us to understand and interpret what the data is saying.

Krall: The key word there is "derive." When talking about the underlying data and its derivation of information, as with options, the underlying is always key. But the second and third derivative of that data is where the edge is. It's taking the "what," and looking behind at the "why" and the "how," and trying to understand how a particular market event has come together.

Big data is extremely interesting and powerful; but big data on its own isn't enough. You need the tools to turn big data into information and intelligence, and tools that are sufficiently user-friendly to do so—that allow users to interrogate data, and present visual market trends and information.

Honoré: I'm not sure I'd weight either as more important, and I'd add a third component. You have to have a curious person leveraging both with a constant sense of "I wonder...." The challenge here is that the barrier to entry is so high. Market data is expensive, so you can't test a new idea unless you've already proven an old idea, because you won't have the capital. When I worked for Nasdaq, I could test anything because I had access. Now, if I have an idea as an individual, access to the historical data and a technology stack is a challenge. This is an innovation barrier I am actively working to break down, and there are a significant number of other people working to do likewise. The other thing I'd say is that we've made the capital markets' structure so convoluted and difficult that it has chased a significant portion of people away. I like companies like Kensho because they're not just combining the data and technology; they're trying to take something a quant does and make it easy enough for an average retail investor to understand.

Adam Honoré
MarketsTech

Maclachlan: Equities will always be sensitive to newsflow. Analytics will not replace conventional newsflow/data as you cannot model news data as a static function. Rumors, announcements and expectations result in uncertain outcomes, if any. Analytics themselves could be considered an additional source of data/newsflow.

Alam: Data will continue to be the driving force behind the power of analytics, with each data source providing a valuable layer of insight for all consumers of analytics. The volume of data sources continues to increase rapidly, adding extra context and meaning to analytical analysis, thus providing consumers with even more valuable insights. The ultimate value of data analytics will continue to depend on the underlying data that feeds it.

IMD: What effect will "app stores" of third-party services operated by data vendors have on niche analytics providers' ability to achieve broader market penetration, and on end-users' ability to source collections of niche analytics alongside best-of-breed data services? Are there any drawbacks to this model?

Rowady: At Tabb Group, we have done some (as-yet unpublished) analysis on the concept of an "Über App Store," where the standards of each applet are sufficiently more advanced than they are today, to allow end-users to develop more complex workflow solutions out of the individual widgets.

In general, yes, the app store concept offers niche players an opportunity for mass distribution. But, because of this ease, the signal-to-noise ratio may stay the same as today. In other words, due to cheaper and easier mass distribution, more entrepreneurs will be able to compete, thereby requiring scoring and ranking mechanisms to help end-users identify the best solutions—the "signals"—more easily.

That issue aside, it is entirely conceivable—if not probable, given additional advancements in standards—that start-up widgets and incumbent widgets (and proprietary widgets, too) could be seamlessly bolted together to form hybrid solutions for complex workflows which deliver dramatically enhanced productivity at much more favorable costs. I think we are a long way from this scenario, but definitely heading in this direction.

Plummer: Certainly the build-out of these data vendor app stores will provide their customers with new and better ways to make use of the information they're acquiring from those vendors. That said, it's important to keep in mind that data variety is a key aspect of analysis. In the face of competition from data vendors' app stores, next-generation analytics providers can, and likely will, retain their competitive edge by offering not only a particular industry or segment expertise, but also the ability to do new types of analytics on new types and sources of data beyond those offered by the established data vendors. Also, who is to say these next-generation analytics providers couldn't white-label their offerings and help build the apps for the larger data vendors? Again, though, it is really about offering users a variety of options and data sources to work from to quickly get them the answers they need.

Alam: The development and evolution of the app store model in financial markets empowers market participants to use the data tools currently available on their market data tool of choice alongside those niche tools they may need for their specific area of focus. Providing market participants with the choice of those tools

they need for their specific role benefits them in a very important way, allowing them to focus on finding the information they need without disrupting their workflow.

Maclachlan: Data vendors will assist niche providers with reaching a larger audience while retaining themselves as the vehicle for the end-user. A possible drawback could be that analytics become commoditized, and deliver less impact.

IMD: How do you envisage the concept and reality of analytics changing in future—both in terms of the types of analyses and underlying data, and also in terms of how they will be delivered? How will the importance of analytics change, and what will be the key drivers of change in future?

Honoré: Here's the future: This isn't some analyst prediction, just fact that has not occurred yet. This is going to be so much easier, and this is where cloud and your previous question come into play. The cloud providers know they're sitting on a gold mine of commercial data, but they don't know what it is or how to bring it out. Their app stores will change that. Go to the Amazon Web Services Marketplace and you'll see you can rent SAP Hana for $0.99 per hour to load all that data into. Grab yourself an instance of Tableau, or something to throw on top of it, and magic starts to happen. We're so very close to that. I would also suggest as this becomes cheaper and easier, you'll see retail people not just come back to the market, but come back to markets that have not been traditional retail plays, like futures and commodities, because the data will give them a lot more confidence and will be easier to understand because technology will make the interfaces so much easier to understand. In talking to start-ups, one of the common challenges they have is that when they create a clean, simple user interface, people who are used to a Bloomberg terminal with 8,000 pieces of information in front of them think it lacks sophistication, when the exact opposite is true. In terms of the data, aren't you more likely to believe in something you discover yourself versus having it handed to you by someone in an industry you probably trust about as much as the average politician?

Krall: Time is precious, and users want to access information and analytics, as and when they need them and in a way that is convenient to them at that time. This will continue to evolve as workflows and tools evolve, and we are making it our business to work with the market around these needs. Today, as it stands, sometimes users of our data and analytics want the ability to interrogate and interact with a platform; at other times, they just want the data served up as quickly as possible, in the most easily digestible format. And they want to be able to specify what that format is, and get it into their inbox as quickly as possible. The ability to offer both interactivity and ease of delivery is key.

Maclachlan: Analytics are likely to evolve further as competition among providers will drive differentiation in search of funkier data. Execution management system providers are likely to include be-

spoke analytics as a standard "extra." This will further educate any untapped audience. Analytics is likely to become commoditized and less of a premium offering.

Plummer: From the perspective of capital markets, the volume and variety of data will continue to increase, as will the pace of analysis. In order for firms to gain competitive advantage, they will need to be able to analyze any and all relevant data available, and make decisions based on that analysis in real time. We see next-generation analytics based on real-time visual data discovery as the way to get there. This capability will quickly bring together large volumes of real-time data with a variety of legacy data, and offer it to users in easily digestible visual representations that will help guide rapid responses to inquiries. We are already seeing the importance of analytics changing: as data becomes more readily available and quickly accessible, the more valuable it becomes. Analytics will increasingly become more important to a broader audience of users.

Alam: With the growing volumes, velocity, and variety of data we are seeing today, it's no longer enough for financial services firms to limit their analysis to traditional market data. To unlock the real benefits of Big Data, market participants need to analyze broader sources, such as unstructured data, and combine that information with existing signals to differentiate and enhance trading, investment, and risk models. The combination of these various data sources and analytics tools empowers financial markets professionals to make more informed trading decisions.

Rowady: Technically, this is all about "AugCog" (augmented cognition). So whatever results in the greatest amount of information conversion will win the day. To head in that general direction, a few things will happen: Raw source data will grow like wildfire. Intermediaries will have a difficult time being disintermediated, as more and more source data is crowd-sourced and contributed. Sensor data will join the pantheon of source categories as automated data collection from network nodes is contributed automatically. For every new question that can be devised, a new analytic will be developed to answer that question. And the whole jalopy will eventually become real-time and on-demand, so end-users can get answers to their particular suite of questions every 4.5 microseconds (or whenever they feel like).

Paul Rowady
Tabb Group

New, never-before considered questions will become the key driver of new analytics, and the accelerating inflation of the value of your time will require that the answer be delivered in highly intuitive and immersive displays (so that the answer to your next question is merely a swipe or a click away). Should be a fun ride! ■

SPONSOR'S STATEMENT

An Added Dimension to Commodity Analytics

Traders and risk managers in physical commodity markets face a range of challenges around transparency, data availability and data quality that create profound problems for advanced analytics. Sam Sivothayan, global product manager at Platts, explains how solutions are emerging to help tackle these challenges.

When it comes to advanced analytics, the physical commodity markets have trailed both their paper equivalents and wider financial markets. Up to this point, it has been almost impossible to carry out detailed analysis of the physical markets, given the complexity and idiosyncrasies of the underlying data.

Take crude oil cargoes: a price quote may appear way off the current market, but if you look behind that price, the specifications or terms and conditions within that contract might explain the apparent discrepancy. For example, contractual terms such as non-standard pricing dates, or Bill of Lading event-based pricing are not typically visible on a cargo bid or offer. Traditional analysis platforms have struggled to handle these complexities.

Something that is simple in other markets—such as analyzing the price action in a particular instrument, including not just executed trades but also the best and worst bids and by whom they were made—is a major task for most participants. Part of the challenge is transparency: because the price action takes place over-the-counter, the data is not easily accessible. And the fact that the price is only one part of the picture adds to the challenge.

The data does exist, but it is usually necessary for market participants to scrape it from trading platforms or information vendors and enter this data manually into their own solutions. Such a process is open to human error, presenting a number of obvious risks that raise concerns for both trade management and compliance. Trading firms are operating in an increasingly onerous regulatory environment, with Dodd-Frank and EMIR (the European Market Infrastructure Regulation) requirements around commodity derivatives disclosure placing ever-greater burdens on the compliance function, leaving no room for avoidable mistakes.

Scraping data also has implications for data quality and, in turn, implications for the quality of analysis. To start with, real-time data collection means that the extent of a participant's historical

> **Something that is simple in other markets—such as analyzing the price action in a particular instrument, including not just executed trades but also the best and worst bids and by whom they were made—is a major task for most participants.**

data tends to be limited to their active involvement in a market—meaning that they will lack the data context upon which analytics depend.

In addition, opaque, over-the-counter environments such as physical commodity markets tend to see revisions made to data after the fact. For example, new participants in Platts' eWindow system are able to transact before they are fully integrated into the system, meaning that the full counterparty information is only entered subsequently. Capturing those corrections and cleaning the data is at best expensive, and at worst impossible. Analytical insights are only as good as the underlying data on which they rely.

There is an answer, however. Energy price reporting agency Platts already captures all the bids, offers and trades across the full spectrum of petroleum contracts submitted to Platts eWindow, including the time, volume and legal entity involved. And not only are headline details captured: for cargo markets, for example, the combination of fixed and indexed pricing used is recorded.

Building an analytical platform that combines Platts' proprietary market data with the understanding to present it in such a way as to transform it from raw data into information and market intelligence was the natural next step. The resulting platform, Platts 4D, promises an extra dimension of market data.

P4D is web-based, user-configurable and provides scrubbed data in preconfigured views, within one hour of the Platts Market on Close assessment process. The platform allows users to see both the data and the patterns within the data, bringing—through insights and advanced analytics—the same trading edge to physical energy markets that other markets have been able to take advantage of.

This kind of tool levels the playing field. It can give traders without deep physical presence in the market the same insights and edge enjoyed by their competitors on the ground. Every trader knows that successful trading is about piecing together the jigsaw from the information they have at their disposal; this tool provides many more pieces of that jigsaw.

In an increasingly time-poor, risk-adverse, competitive environment, traders, their analysts, risk managers and compliance can at least breathe a sigh of relief and cross one task off their list: P4D is the first off-the-shelf solution to deliver robust, reliable, transparent and insightful analytics to the physical markets, and it promises to be a game-changer. ∎