# waterstechnology

# After Bots Learn to Read, They Learn to Write

ADVANCEMENTS IN NATURAL LANGUAGE PROCESSING ARE FUELING ADOPTION OF NATURAL LANGUAGE GENERATION. WHILE THERE'S GREAT PROMISE IN THE FIELD OF NLG, THE BARRIERS TO ENTRY ARE MANY.

# waterstechnology

# Customise your email alerts with Follow Topics

Our most popular
Follow Topics include:

- Data management
- Artificial intelligence
- Management & Strategy
- Alternative data

# The Best-Laid Plans

**Innovation is just as** exciting as it is horrifying. For example, as social networks and algorithms grow more sophisticated, I am terrified that my social media scores will eventually determine how I live, like Bryce Dallas Howard's character in *Black Mirror*'s "Nosedive" episode. Although I was a painfully awkward teenager and sometimes an absolute tool—and some might say I'm still those things—I'm happy that I grew up in a simpler time before the internet permeated all facets of life and when caller ID was considered an advanced phone feature.

But rather than fret about dystopian futures and *Black Mirror*-like events, let's focus on the positives—specifically, natural language processing (NLP). This year has been the year of NLP and many hope that soon the spotlight will shift to natural language generation (NLG).

Jo Gallagher wrote a deep-dive feature on how NLG is slowly seeping into the capital markets and what the challenges are (*see page 22*). Banks like the Royal Bank of Canada and Morgan Stanley, as well as data giants like Bloomberg and Refinitiv, are experimenting with—and even rolling out—NLG tools. While there is a ton of promise in this field of artificial intelligence (AI), as usual, it's not a plug-and-play tech in the capital markets because of how highly regulated financial services firms are.

There are some very good use-cases in Jo's story, but several sources were also careful to note that NLG is not something they will rush to unleash on a bank or asset manager, especially when NLG is combined with machine learning. NLG will have its day, but that day is farther off than some might claim, at least in the capital markets. The same could be said about Reb Natale's feature on agent-based modeling (*see page 18*).

Innovation is equal parts vital and dangerous. That's where the testers come in. This month, Max Bowie spoke with a swath of technology providers to get their thoughts on best practices when it comes to testing (*see page 12*). Before a company can release a new technology to clients, it must be rigorously tested. In fact, many firms will do everything in their power to break shiny new toys that so many engineers, data scientists, project managers, and testers have worked so hard to create.

Technologists are supposed to move fast and break things—and it's because of this that we've seen incredible tech breakthroughs come from the likes of Facebook, Google, Amazon, and Twitter. Similarly, we also now have the cesspools created by the likes of Facebook, Google, Amazon, and Twitter. If NLG or agent-based modeling is to take hold in the capital markets, it will be in large part thanks to the testers. But as with any technological revolution, there's an inevitable, unforeseen backdraft that even the testers won't be able to find. Let's just hope this results in awkward growing pains instead of cataclysmic market shifts. **wt**

**Anthony Malakian**
Editor-in-Chief

**waterstechnology**

# Contents

**GET MORE ONLINE**
**News. Webcasts. Video. Audio. Special Reports. Get it all at waterstechnology.com**

# Fidelity Shutters ActionsXchange; Corporate Actions Biz Margins Blamed

Officials remain silent on the closure, which has been in the works since last year, sources say.
By Max Bowie

Fidelity Investments is in the process of shutting down its Fidelity Corporate Actions Solutions (formerly known as Fidelity ActionsXchange) business line and its corporate actions data validation service, prompting an exodus of clients to alternative solutions, *WatersTechnology* has learned.

The firm decided to bow out of the space in August 2019, sources say, and clients have spent the past year identifying and migrating to alternative providers. The vendor informed clients but has not made any formal announcement about the closure or given any reason for its decision. And those providers tasked with filling the gap left by Fidelity's departure report drastically differing experiences: Some vendors say the move has handed them an influx of new clients, and that Fidelity collaborated with clients and vendors to identify alternative solutions, though others say Fidelity has provided little assistance in transferring clients to new providers.

"Customers are very angry. They've been left in the lurch. Some of them had long-term contracts," says an executive at one data provider, who says it was months before Fidelity informed them of the decision, and took even longer to transition clients because Fidelity—not the corporate actions data providers themselves—controlled the relationships with the end clients of its service, making it difficult to ensure a smooth transition.

Fidelity Corporate Actions Solutions provides a cleansed and normalized managed feed of corporate actions and tools for handling the corporate actions process at end-user



> " The cost dynamics of running a managed service in this space are pretty tough to manage. Costs are often higher than revenue, and you need global critical mass."
> **Virginie O'Shea, Firebrand Research**

firms. It captures corporate actions feeds from third-party vendors, then formats the announcements into a composite record. It also combines validated information with data from clients or their custodians into a composite summary that clients can use to monitor key event dates and manage exceptions via a workstation. Additionally, it offers a notification and response workflow tool for automating instructions to portfolio managers and custodians, and software to calculate and automate the processing of payments arising from corporate actions.

Sources say two factors most likely prompted Fidelity's decision to close the business: First, other vendors have made inroads into this market in recent years, and second, the nature of corporate actions processing as a low-margin

business. The costs of operating in this space can be high—particularly the cost of acquiring corporate actions datafeeds from multiple vendors before even applying any proprietary value—and as the oldest provider in the space, Fidelity may have needed to make technology upgrades that simply weren't worth the expense.

"The cost dynamics of running a managed service in this space are pretty tough to manage. Costs are often higher than revenue, and you need global critical mass," says Virginie O'Shea, CEO of Firebrand Research, whose first research paper, published this summer, provided an analysis of the corporate actions space.

### Replacements

Sources say the most likely replacements are IHS Markit's IMActions and FIS' XSP services, though data management platform vendor GoldenSource is also reporting positive interest in its services, despite not being a direct competitor.

"Fidelity ActionsXchange is a long-standing and important participant in this marketplace," says John Eley, CEO of GoldenSource. "Their decision to alter their business model has highlighted the need for software and service solutions that are flexible and can work with any data provider. We've already landed a client as a result of Fidelity's actions, and are in talks with others about providing a solution that's flexible in terms of workflow and support for other data sources."

It's not clear when Fidelity will stop supporting clients. Fidelity officials did not respond to requests for comment. **wt**

# AML Platforms 'Are Just Not Working' —Is Human Error or Tech the Culprit?

Technologies meant to prevent financial crime may add more headaches than they relieve. Others say the best tech can't overcome how bankers use these tools. By Rebecca Natale

Despite the introduction of machine learning, natural language processing, and the burgeoning field of regtech tools to help firms fight financial crime, banks are still falling victim to—and perpetuating—money laundering, a *de facto* economy in its own right that amounts to an estimated $800 billion to $2 trillion globally, according to statistics by the United Nations' Office on Drugs and Crime.

Legal and tech experts attribute this to a disjointed approach to compliance technology across the board, from the range of third-party vendors available, to a smattering of proprietary systems and legacy technologies, to poor implementations of what could, in theory, be otherwise-working systems for flagging and stopping illegal activity.

"If you look at some of the transaction monitoring systems of most of these larger banks, it's usually like a bespoke sandwich of proprietary things, third-party things—and it's kind of a mess," says Spencer Schulten, US head of financial crimes compliance at Capco. "And when you don't have your threshold tuned, or you don't have people properly conducting an investigation with a full knowledge of what they're doing, you might just have a lot of suspicious activity reports (SARs) that don't need to be filed."

SARs are reports filed by financial institutions about suspicious or potentially suspicious activities to the Financial Crimes Enforcement Network (FinCEN). In September, SARs served as the basis for a massive data leak that has come to be known as the FinCEN Files. The files revealed more than $2 trillion in transactions between 1999 and 2017 that were flagged as potential money-laundering schemes, yet often little or no action was taken by the filing banks or FinCEN, and just as often, the transactions were facilitated unhindered.

Four days before the files were published, FinCEN sought public comment on how the organization should enhance the effectiveness of its anti-money laundering (AML) programs. Indeed, Schulten wrote in October that the files show systemic failures in the prevention of financial crimes.

Though the liability for such failures ultimately falls on the banks, institutions employ platforms and services from a variety of tech vendors to help them manage the deluge of transactions facilitated every day.

"You hear a lot of rumblings in the marketplace because the single-point solutions, and the solutions that are out there today, are just not working," says Melissa Townsley-Solis, CEO and co-founder of GIACT, a company that specializes in digital identity, payments verification, and fraud prevention, and which was recently acquired by Refinitiv's risk business.

"You're seeing fraud increase by 15% in some industries. Some industries even higher than that," she says. "Obviously you look at the US federal government, which has struggled to even pay out stimulus payments to the right people … so it's now caused people to look at their solutions and say, 'We have to do something different because [the] status quo doesn't work anymore.'"

However, that isn't to say they're not working outright. The deeper problem may be that they're not functioning as
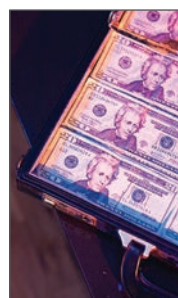
AML systems can be ineffective and poorly implemented

well as they could, and that is a problem that sits with the client as much as the provider, says James Mirfin, head of digital identity and financial crime at Refinitiv.

"You can take the best technology and implement it badly, and you're not going to get the outcome that you expect," Mirfin says. "Unfortunately, [as] you see companies run toward digital and digitization, you can see some bad implementations of either where they put too much control in place and it impacts the consumer experience, [or they don't put enough control in place]. It's a balance between friction and protection in many ways, but it is possible to get this right."

It's the perennial problem in which technology is created to make some aspect of life easier, but in turn, makes others harder. So goes the story of false positives and negatives in transaction monitoring, a side effect of machine learning and artificial intelligence. In 2018, Reuters reported that as many as 95% of system-generated alerts are closed as "false positives" in the first phase of review with about 98% of alerts never culminating in a SAR.

Matthew Banham, a partner who specializes in financial crime and regulatory enforcement investigations at global law firm Dechert, boils down the central challenge to time and value.

"Is there resource and time within the business to do rigorous testing as to who your customer is, why they're doing that transaction, [and] what it's about? And if you can be that inquisitive, can you do it in a way that's not going to offend the client, or cause them to take their business elsewhere?" Banham says. wt

# Nasdaq, HKEx: Blockchain's Disruptive Potential is Overstated

Distributed ledger technology won't replace existing exchange infrastructure despite potential for smaller use cases, say CEOs. By Wei-Shen Wong

The potential for blockchain technology to disrupt the exchange market is a "bit of an overstatement", the heads of two major exchanges have said.

In a virtual discussion at the Hong Kong FinTech Week 2020, Charles Li, chief executive of Hong Kong Exchanges and Clearing (HKEx), and Adena Friedman, president and CEO of Nasdaq, agreed that blockchain has been overhyped and that the decentralization of exchanges is unlikely.

The two executives said the idea that distributed ledger technologies could "change everything in the exchange world," as Friedman put it, has been exaggerated. "There are a lot of practicalities to implementing blockchain that make it a long-term evolution [rather than a short-term disruption]," she said.

Friedman believes that while blockchain can create better and immutable records, it isn't a panacea for all operational inefficiencies and will not render intermediaries irrelevant by decentralizing markets. However, the technology would be a good grounding for a new market built from scratch, she said, as it provides an efficient structure from trading through to settlement.

Li said exchanges are unable to progress unilaterally with innovations and technological developments because they perform the critical function of a centralized market. They must therefore operate with consensus and involve market participants at every step of the way, he said.

"I used to say that an exchange is like a pack of wolves moving in snowy mountains. The stronger guys [market participants] are at the front, and the

Blockchain is no panacea for all operational inefficiencies

exchange is the strong wolf at the very end, because we have to move the slower members along. We can't just abandon them in the field."

However, he added, blockchain technology does offer the potential for innovation in non-core operations and processes that the exchange or market participants don't rely on day to day.

"It's very hard to change the core operations when everybody else is still using a particular older technology. You can't really force everybody to change overnight," he said.

Li added that for new initiatives or additions to HKEx's services, such as its Stock Connect programs, the exchange has started experimenting with using blockchain. "[This is to see] whether we can do the new reporting so we can comply with the Chinese onshore rules more effectively. That seems to be a welcome change, but we are still on the journey of trying to see how to make it happen," he said.

Nasdaq's Friedman said that a complex web of participants are involved with incumbent exchanges in the highly regulated securities markets, and it's only natural that some are more technologically advanced than others.

## CHESS Moves

"Also, there are all these offline systems that record all of the information back into the banks' and asset managers' systems. You have to bring that full ecosystem along in order for the blockchain to really make that change. That is a long journey, as we've seen in Australia. It takes a long time to change from traditional technologies to something new, and at the same time you have to say, 'Well, what is the benefit

that the end-users are getting?'"

The Australian Securities Exchange (ASX) has been working since the end of 2017 to replace its equities clearing and settlement system (CHESS) with a blockchain or DLT-enabled platform developed by Digital Asset. It was initially set to go live by the end of this year, but was later put back to April 2021.

Then, after Covid-19 hit, the exchange decided to review its implementation timetable to allow market participants to focus on their day-to-day operations. In June, it set April 2022 as the go-live date to replace CHESS. After consulting with the industry, ASX said that as of August 4, 91% of CHESS users who made submissions could meet the revised date.

Since then, however, the exchange has delayed the launch until April 2023, saying many CHESS users have asked for extra industry testing and more time to prepare.

As time has passed, many in the industry have soured on the promises of blockchain. Bill Murphy, long-time CTO of Blackstone, who left in March and joined Cresting Wave this summer, recently said on a *Waters Wavelength* podcast that he still has not seen good use of blockchain outside bitcoin. "It's a solution searching for a problem. I think we will continue to see lots of noise, but no real results from blockchain," he said.

At the end of the day, said Murphy, blockchain is just a database: "The only reason you need blockchain is when you have truly anonymous transactions that need to be trusted in a way that can enable that. There aren't that many of those use cases in the world." wt

# Northern Trust Puts Faith in DLT for the Future

Distributed-ledger technology will help ensure asset safety and allow custodians to provide more value to clients using AI and analytics. By Wei-Shen Wong

Though the allure of distributed-ledger technology has waned for large-scale use within the capital markets, some market participants, such as Northern Trust, are still investing in the technology because execs there believe it has the potential to offer valuable change over the next decade for targeted use-cases.

In a whitepaper published in October reimagining what custody will look like in 10 years, Northern Trust details how over the past decade, global custodians have moved away from their traditional roles and begun providing additional services around cybersecurity, artificial intelligence, cloud computing, data analytics, and blockchain development. It predicts that in the coming decade, technologies such as DLT will be crucial to custodians as the issuance of digital securities increases and new settlement platforms emerge.

"By 2030, custodians will be providing a new, collaborative ecosystem where both digital and traditional electronic solutions exist side-by-side. The custodians of 2030 will be client-centric and focused on asset safety. But they will also be flexible, agile, creative, and digital in ways we haven't seen before," the firm writes in the whitepaper.

The world of securities services has evolved from when buying a stock in a company involved the delivery of physical certificates to prove that you own some part of that company to a point where certificates—and even the securities themselves (such as 2018's world-first blockchain bond, issued by the World Bank and Commonwealth Bank of Australia)—are now digital.

One of Northern Trust's previous efforts using DLT was a solution for private equity firms to open, manage, and administer funds. It has since transferred this DLT platform for private equity asset servicing to Broadridge to develop further. It is also working with Singapore-based BondEvalue to deliver integrated asset-servicing and digital solutions for fractional ownership of fixed-income bonds.

Danielle Henderson-Gerace, head of market advocacy and innovation research for Asia-Pacific at Northern Trust, says moving from electronic to digital will change how assets are issued, exchanged, and settled. This could lead to entirely new digital securities across all asset classes, including bonds, private equity, exchange-traded funds, special-purpose vehicles, precious metals, real estate, and others.

"Distributed ledger is part of that story and the flexibility of what you can do with that digitalization of asset classes. That has required custodians to think about the future and have a pathway to develop capability and active experimentation, and give our clients confidence that we're going to be ready when that's our new normal," she says.

Henderson-Gerace, who leads a team focused on researching, incubating, and experimenting with various technologies so Northern Trust can bring real-world solutions to its clients, believes that within 10 years, these new digital asset classes and market infrastructures will co-exist—at least for a period of time—alongside those traditional electronic marketplaces.

"If you think about what DLT enables, it is the data, the processes, the lifecycle management, and efficiency. But as we said, the big shift in that sort

DLT will prove crucial to custodians, says Northern Trust

of custody reimagined is not just [that] we keep it safe, we process, we administer. … Technology is going to play a greater role and change the way we offer expertise in security servicing," she says.

When asset safety is ensured, and data is transparent, distributed and accessible, the custodian's job would be to use analytics and AI to create personalized experiences for how investment managers access information, and to support actionable insights.

An example that Northern Trust is practicing internally is using cognitive document analysis for unstructured data. Henderson-Gerace says this process has been applied to legal agreements, and is particularly useful when looking at inconsistent clauses covering one particular aspect of regulatory management.

"You can train software now to analyze and identify common clauses. Lawyers can navigate and train the machine to find that specific clause, so what was historically a three-month job among seven lawyers can be essentially a two-week process. … This unstructured data could be in documents generated in private funds where there's not a lot of consistency. So, how can we use AI to help us drive both extraction and centralization of that data in a more efficient way?" she adds.

Henderson-Gerace declines to provide details of the projects Northern Trust is currently working on, due to confidentiality issues, but she says the general themes are around making new digital asset classes available in existing markets and systems, and bringing them up to regulatory and institutional-grade standards. wt

# EU Firms Seek Clarity on Tech Rules for
## Incoming Cloud Guidelines

As Europe prepares for a new batch of outsourcing rules, some firms are looking for answers on how to test exit strategies and mitigate concentration risk. By Jo Wright

Ahead of the European Securities and Markets Authority's (Esma's) upcoming rules on cloud outsourcing, questions surrounding two issues—testing exit strategies and managing concentration risk—are stifling firms' preparations. While Esma is set to publish its final guidelines on cloud usage early in 2021, some industry firms are urging the EU regulator to provide further clarification on how to comply with the rules and prepare their tech stacks.

In responses to the Esma consultation, seen by *WatersTechnology*, but yet to be published, several industry firms have highlighted key operational concerns. In one submission, a European market structure firm writes that cloud exit strategies could amount to significant technical spending that require extensive code rewrites and retesting of operations moved to the cloud.

The response adds that testing exit strategies in some cases may not even be possible due to the efforts required to migrate data and applications from one location to another—for instance, from a cloud service provider (CSP) to on-premise or to another CSP.

"It's not simply a copy and paste exercise," says a senior executive for regulatory policy at an investment bank, echoing the market structure firm's concern regarding exit strategies. "It depends on the type of cloud service we're using, and it can be much more [work], especially, when we're not just talking about data storage, but co-innovation, [co-development] or microservices like API-as-a-service."

The Esma consultation outlines that firms "should develop and implement exit plans that are comprehensive,

New guidelines for cloud usage in Europe are expected in 2021

documented, and sufficiently tested." The draft says regulated institutions could be expected to develop transition plans that include "trigger events" that would activate an exit strategy and test them using a risk-based approach.

In a second response to the consultation paper, another market structure firm calls for clearer guidelines on what the testing requirements entail. In its submission, it writes that rather than physically testing systems, a firm's exit plan can be tested by reviewing documentation on areas such as co-dependencies, alternative service providers, the CSP's contractual obligations to support the strategy, and the firm's resources for executing it.

When it comes to cloud setups, there is no one-size-fits-all. Each firm manages its data differently. The same applies to testing cloud arrangements, says Douglas Wilbert, managing director of risk and compliance at Protiviti. "The concept of an exit strategy and testing is something that is on the organization and how they set up their cloud strategy," he says. "And if they cannot test an exit strategy [when there is] a failure, they are probably going to be in trouble."

Testing cloud arrangements will not only involve the firm's CSPs and on-premise locations, but their third-party solutions that operate on the cloud. Therefore, financial firms need to review their vendor's cloud concentration and how that impacts their operational resiliency.

The Esma consultation on cloud outsourcing is one of several regulatory guidelines being published or proposed. In all of the responses to the Esma guidelines seen by *WatersTechnology*,

firms have sought further harmonization of global outsourcing and third-party rules.

In financial services, there are two types of concentration risk: the firm's individual concentration risk and sectoral concentration risk. In the first response to the Esma guidelines, the market structure firm outlines that the latter risk should be primarily countered by the regulator, rather than individual firms, as firms do not have the oversight or the authority to influence their competitors on what CSP they should use.

"The other thing is that, in order to identify where there is concentration risk, a lot of the underlying data required is simply not available to us. So it's not just that a financial institution is not able to tell another institution what to do or not to do, it's that we simply wouldn't even know which CSP is contracting with which of our competitors," the senior bank executive says.

Similarly, in feedback submitted by the World Federation of Exchanges, the global trade association for exchanges and clearing houses, the group said, "It is not clear how in practical terms such a firm/group could ascertain how many of its peers are also using the same CSP," and, "CSPs are typically restricted by contract from the disclosure of other firms' use of the services."

To avoid individual concentration risk on one CSP, some regulators endorse the adoption of hybrid or multi-cloud strategies. The counter to this view is that one of the drivers for firms to move to the cloud is to simplify their technology stack and offload complex legacy systems, but if they are mandated to use multiple providers this could offset the original objective. **wt**

# Asset Managers Fear ESG Data Disclosure Gap

Investment firms need data to meet upcoming regulatory requirements. But corporates aren't making this data available in high enough quantity or quality. By Jo Wright

**S**tarting in March 2021, asset managers in the EU will have to publicly disclose information related to the sustainability impacts of their products and of the companies in their portfolios so that investors can make informed decisions about environmental, social, and governance (ESG) investing. To meet the so-called disclosure regulation requirements and other data-heavy ESG rules coming out of the EU, these investment firms will need reliable, comparable data from the corporates in their portfolios.

But corporates aren't compelled to report ESG information in anything like the granularity and specificity required by these buy-side firms—which now fear a costly "data gap" in the information available to them.

"What we need from these industry companies in the future is a pretty deep understanding in terms of the environmental impacts of [their] activities," said Jarkko Syyrilä, head of public affairs at Nordea Wealth Management, speaking on a panel during the TSAM Digital: Regulations and Compliance event.

Corporates are currently required to publish information on their policies and business models related to concerns like the environment or corruption under the Non-Financial Reporting Directive (NFRD). Buy-side firms, however, say this reported information is not enough to populate the detailed disclosure templates they will have to post on their websites. Not only is it insufficient in quantity and quality, but it's also not standardized, and therefore neither reliable nor comparable.

The EU supervisory authorities earlier this year consulted on the regulatory technical standards (RTS)
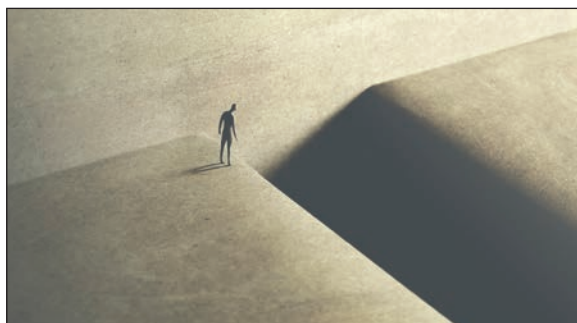


Mind the gap: ESG information provided by corporates is often neither reliable nor comparable

of the disclosure regulation. The RTS will probably only be finalized weeks before firms must begin to report, but it seems clear that regulated firms will be expected to calculate the adverse environmental impacts of products or companies using a set of indicators, such as carbon emissions and deforestation.

"Those kinds of indicators that have been suggested are going to deeply impact these companies. Clearly, we need to understand issues, like what is the energy usage of those companies. But then going further to human rights issues, all sorts of issues—that is over 50 indicators—is quite a data-heavy exercise. And currently that data doesn't exist: Listed companies don't have obligations to publish this data, so it is very sporadic," Syyrilä said.

"We don't know what the final indicators will be, or what level of disclosure will be required from asset managers, but we can expect them to be very detailed, and over time the detailed requirements will only grow."

The EU adopted the disclosure regulation—called the Sustainable Finance Disclosure Regulation, or SFDR—in late 2019, and the related Taxonomy

Regulation in June. The disclosure regulation requires investment firms to display on their websites public disclosures about the sustainability risks of their products, and the companies in their portfolios. To make these disclosures, these firms must collect the data and calculate the sustainability risks and impact using a set of indicators.

The European Commission has acknowledged that corporates don't report enough data to match the evolving needs of asset managers, and consulted on a review of the NFRD to bring the reporting of listed companies in line with the disclosures the EC is requiring of asset managers.

Even so, given that the NFRD review was only proposed in February, improvements to it will only come into force a year, or even two, after the disclosure regulation's requirements kick in. The EU has already adopted the disclosure regulation, which takes effect on March 21, 2021. The European Fund and Asset Management Association says in feedback to regulators that even if the NFRD review goes swiftly, "the disclosures cannot be reasonably expected before 2024."

"There will be a gap of one or two years before there are actually obligations for those industry companies to publish this data, which we will need to be able to fulfill the obligations on us as asset managers. So it is a bit of a challenge," Syyrilä said.

In the absence of reliable data from the corporates, asset managers are considering asking third-party providers to plug that data gap. "It's going to be a gold mine for those data vendors, of course. We are paying dearly for this data," Syyrilä said. **wt**

# Wells Fargo's Head of Quant Strategy
## Explains HPR Deployment

Wells Fargo's Quantitative Prime Services division has tapped HPR's Unimus platform, starting with its market access gateway and risk management tool. By Anthony Malakian

John Leone has spent almost 30 years working on the buy side, most recently at Matthew Tewksbury's Stevens Capital Management (and Tewksbury Capital Management and Trout Trading Capital Management), where he managed market connectivity, direct market access, and prime brokerage and financing relationships. This is all to say that he knows the needs of a quantitative hedge fund.

In the summer of 2019, he decided to cross the street, joining Wells Fargo as the bank's head of quantitative strategy for its Corporate and Investment Banking unit. When compared with other Wall Street banks, Wells Fargo has not been as aggressive in the quant-tech investment space, but Leone says this offers the group a late-mover advantage. And rather than build, Leone's three decades of experience led him—and thus the bank—to partner with HPR (formerly Hyannis Port Research).

"There's been a lot of smart discussion about how challenging it can be to compete with Silicon Valley for the best engineering talent. With certain segments of the technology stack, in particular what HPR focuses on, that's a Herculean task [to build that from scratch] when you look at how the massive tech companies of the world are able to recruit, and how they build engineering teams from very early stages, and have them evolve organically. Our partnership with HPR helps us level the playing field," Leone tells *WatersTechnology*.

Wells Fargo is now live with eight large hedge funds on the Unimus platform. To start, the bank is incorporating HPR's market access gateway and its

Wells Fargo is aiming for "late-mover" advantage

pre-trade risk management tool on its Quantitative Execution desk, though Leone says the bank will migrate other business units onto the platform in the future, as well as add other services, such as HPR's market data distribution and surveillance offerings.

The key reason for selecting HPR was the flexibility the suite of services provides, Leone says.

"We're not forced to push some off-the-shelf capability on clients, which is what firms who have legacy, piecemeal solutions often have to do," he says. "With HPR, we can go to a startup manager that may not have super-advanced order-routing methodologies and things like that, and we can develop a custom solution. We can also go to a very large, established group that has specific ways they do things, and offer them self-directed DMA, [and a] co-located, FPGA gateway."

Anthony Amicangioli, CEO of HPR, says that of the eight Wells Fargo clients that have been migrated onto the platform, the first two implementations were similar, but the following six needed customization.

"If you're going to build a platform, it really needs to be able to be morphed and flexible based on those [individual client] needs," he says. "So, for example, if you take something as simple as FIX connectivity, a lot of people think of that as being like a USB plug-in, and you just plug that into your computer [and it runs]. Unfortunately, the reality is, I don't think any two FIX connections, as a simple example, are alike; you always have to be able to [customize that connection]. If you can't do that in an agile, predictable, and high-performance way, you're in trouble."

While Wells Fargo has an existing DMA platform, Leone says it wasn't built with an eye toward connectivity for low-latency, quantitative hedge funds that require the cleanest engineering and fastest, most-stable pipes; HPR made its name by developing high-performance FPGAs, allowing the bank to focus on high-touch services.

### One-Stop Shop

As *WatersTechnology* has noted previously, HPR made its name as a hardware provider, but the company has spent the past several years expanding into the software space. With its 2018 launch of Omnibot, which serves as a low-latency switch, router, and pre-trade risk gateway, HPR began to move in a new direction. Omnibot leverages HPR's Unimus management and control framework, which underpins the vendor's entire solution set, including Riskbot, Softbot, and Databot.

While Unimus has been around for many years, HPR is now leveraging it to grow its software presence. The goal being that it wants to be a one-stop-shop for firms of all stripes and sizes, thus making it a more enticing platform for banks like Wells Fargo to incorporate on the quant side, "soup-to-nuts," as Leone puts it.

"While [we offer] market access, risk gateways, pre-trade risk, etcetera, the reality is that the big game in this platform business is the ability to begin to fold all of the products onto a very singular platform … When you look at what Google and Amazon are doing, it's very similar—they are shepherding users into a more singular way of doing things. So the trend is very clear," Amicangioli says. wt

# OPEN OUTCRY

"M&A in this industry is rarely beneficial for customers. The most welcome ones are … where there is not much crossover, such as a data company buying a technology provider. But when one fish swallows another, it just creates more dominance and less end value for clients." **Market data exec at a European asset manager**

"I would be pretty shocked if any large company let one of these things loose on their customers, or in any other specific case, because you really don't know what it's going to do, like it is a probabilistic model. And it is basically just generating the next word based on things it has seen already, so it is quite unpredictable. We are not using anything like that." **Paul Tepper, an executive director in Morgan Stanley's technology division**

"There's been a lot of smart discussion about how challenging it can be to compete with Silicon Valley for the best engineering talent. With certain segments of the technology stack, in particular what HPR focuses on, that's a Herculean task [to build that from scratch] when you look at how the massive tech companies of the world are able to recruit, and how they build engineering teams from very early stages, and have them evolve organically. Our partnership with HPR helps us level the playing field." **John Leone, head of quantitative strategy for Wells Fargo's Corporate and Investment Banking unit**

"All models are ultimately wrong because they aren't the real world. But you want to have an ecosystem—a zoo of models. A zoo of models: a whole bunch of different, diverse species of models, so that you're not blindsided." **Justin Lyon, CEO of Simudyne**

"If you look at companies like Netflix, Facebook, Microsoft, and others, they are probably using or starting to use Chaos Testing. But in the financial industry, firms are probably only just getting started using this." **Derek Ferguson, head of technology in Fitch Group's IT development solutions division**

"There are all these offline systems that record all of the information back into the banks' and asset managers' systems. You have to bring that full ecosystem along in order for the blockchain to really make that change. That is a long journey, as we've seen in Australia. It takes a long time to change from traditional technologies to something new, and at the same time you have to say, 'Well, what is the benefit that the end-users are getting?'" **Adena Friedman, Nasdaq**

"Customers are very angry. They've been left in the lurch. Some of them had long-term contracts." **Executive at a data provider**

"There have been loud cries for transparency in market data pricing because it is considered to be opaque. Every different provider of market data has a different pricing policy, and they can run into 60, 70, 80 pages. So where there is some standardized template for how market data pricing is done, whether that will change anything, I doubt it. And we will probably have a 70- or 80-page template anyway to try and fit in what everybody wants to get in. It doesn't truly address the problem, and it doesn't simplify or make market data more transparent at all, from what I can see." **Graham Dick, CEO of Aquis Exchange**

# Testing, Testing…



Just as a doctor wouldn't administer a vaccine to a population without adequate trials, fintech vendors don't release new products or updates without putting their software through rigorous testing. What does that testing involve, and what different approaches do companies employ? Max Bowie finds out. With additional reporting by Wei-Shen Wong

They ride in the tour bus, but they aren't the rock stars. Behind the scenes at any concert, an army of sound engineers and audio technicians ensure that the music reaches the audience. The capital markets also have rock stars of their own—the high-profile traders whose activities make headlines, and the technologists who build the systems that can trade off of headlines.

Capital markets firms and their technology providers take the rollout of new software as seriously as a musician takes their sound quality, employing organizations devoted to testing and re-testing code before being deployed into production, and leveraging any and all new tools and techniques to put code through its paces—even if it means breaking it and having to start again—to ensure any new

software is battle-tested before being deployed to the front line.

They take it seriously because of what's at stake: money—and lots of it. "Our products are used to trade or manage risk and compliance in the financial markets, so by their nature, any releases that we do are always very high risk," says Jim Nevotti, president of trading and risk management software vendor Sterling Trading Tech. "If we release a trading platform and get the symbology or something else wrong, or if it doesn't support a certain order type, there are large amounts of money associated with what can go wrong."

Errors resulting from a lack of testing can be costly—not just the risk of clients

losing money if something doesn't work correctly, but also the time and money required to investigate problems, fix software, and roll out patches, as well as any lost revenue from dissatisfied clients.

But before investing time and money in building software—let alone testing it to the breaking point—organizations should first test their hypothesis: Should they really be building this particular product or enhancement? Is it really necessary? And just because it's a cool idea, or a few clients ask for it, will that translate into something that provides sufficient value that users will be willing to pay for it?

"First, we assertively work to validate that what we are doing has value for the

customer," says Terry Roche, co-founder and CEO of startup enterprise technology vendor Pegasus Enterprise Solutions. "We attack it from every angle, and make sure it is entirely validated with clients applying our deep, real-world experiences. A lot of organizations don't do that."

This is an important strategy when approaching "moonshot" projects that go beyond business-as-usual tasks and routine process automation, says Agrim Singh, hacker-in-residence at Citi Ventures D10X, a division of the bank's venture arm tasked with identifying new technologies for use by Citi and its clients.

"Part of that is obviously you want to be doing a lot of generation of ideas and generation of the problem statements you want to tackle. But you also want to find ways of assessing risk around these [ideas and problems by] building and breaking things so that you can quickly validate or invalidate whether your assumptions are correct, your problem statement is correct, whether your solution hypothesis is correct, and there are various steps within the product line," Singh says. "That's part of my role as well—to not only talk to the people at the beginning to validate problems, but also build and break things to see what you receive at the end."

Singh is a fan of moving quickly and testing concepts internally and with clients to prevent wasting time developing and testing ideas that may be flawed or unwanted, and is a proponent of hackathons that can validate ideas within a day, and getting "quick and dirty" demonstration products into clients' hands quickly to solicit feedback.

FactSet Research Systems' recent project to migrate its ticker plant to the cloud is an example of a project that was validated on paper, but where the vendor wasn't entirely convinced of its viability—and whether cloud could offer the stability and low latency required by such a critical piece of infrastructure—until it had put the cloud ticker through an exhaustive series of tests and could be fully confident in its resilience, which proved especially important given this year's volatility.

> "With any defect, prevention is definitely better than a cure. You want to introduce practices that prevent regressions from sneaking into production. The goals are to spot bugs as early as possible in development, and to fix them fast—even when they are spotted late, and even when they are spotted after they are already deployed to customers."
> **Denis Chekhlov, Bloomberg**

"Historically, and this year especially, we've seen some incredible volumes. We were able to simulate … close to 10 times the highest volumes we've ever seen. We could see the impact of that load, and we could see the environment being loaded differently from the average, but there was no impact on clients. You see a drop in terms of latency, but we were still achieving our service-level agreements with plenty of headroom," says Gene Fernandez, chief product and technology officer at FactSet. "Our engineers started out completely skeptical that it could even be done. But then they started to see statistics that showed it might be possible. And then they saw statistics that we might be able to materially improve things. So they went from being skeptics to being advocates."

The vendor spent the first half of this year planning and testing the cloud ticker, stress-testing it and deliberately trying to find ways to make it fail, to understand its limits and learn how it would react when faced with certain situations, from unexpected bursts of market volume to systems outages.

To simulate outages, FactSet employed a tactic known as "gorilla testing"—"like if a gorilla got loose in the datacenter and started ripping out cables and servers—so we can see how the system responds," Fernandez says.

In these simulations, the vendor began by trying to imagine what scenarios might occur. Then it assigned engineers to a "red" team and a "white" team. The red team spent its time doing all it could to take the system down, while the white team observed their efforts and

**Denis Chekhlov**
Bloomberg

**Terry Roche**
Pegasus
Enterprise
Solutions

chronicled the impact on the system and its environment without intervening or doing anything to defend against the red team's efforts.

## Prevention Is Better than a Cure

Many of the things a firm or vendor will test for may not be major faults, but rather changes in an update or new feature that cause previous features to not work properly. The new feature may have been tested independently and found to work fine, but code does not exist in a vacuum. It has to interact with other components and work seamlessly as part of an enterprise-wide software framework.

"Environmental dependencies may not always be explicit as to how a change affects other areas," and software bugs can be code that works as designed, but the designed usage may be "erroneous behavior deliberately introduced as a result of misunderstanding or miscommunication around functionality requirements," says Denis Chekhlov, chair of Bloomberg's automated testing guild. "So at each point, you need to think about what you are testing and for what purpose—the business logic itself, and the underlying integration with other systems and data."

Testing also can't be viewed as a substitute for due diligence during the development process. Compromising or skipping steps during development—or not ensuring that different components that are part of an ecosystem involving multiple teams of engineers and business staff work properly together—may make it harder to test the entire code base once complete.

"With any defect, prevention is definitely better than a cure," Chekhlov says. "You want to introduce practices that prevent regressions from sneaking into production. The goals are to spot bugs as early as possible in development, and to fix them fast—even when they are spotted late, and even when they are spotted after they are already deployed to customers."

And because bugs may not be coding errors but may be something that arises as a result of unintended use, issues can't always be completely prevented, adds Sterling's Nevotti.

> "If you can auto-generate some of the functional tests, you can focus more on business testing, and being able to test the way end users would. For example, you may find usability problems that technical test staff didn't know to test for."
> **Tom Stock, GoldenSource**

"There are always unintended consequences from real-world use of your software, and you can never account for everything," he says. "You just have to find them and modify your procedures to account for them."

But the more a company has the resources to put software through its paces internally, the less chance there is that any bugs will wind up in customers' hands. Caleb Eplett, chief product officer at fundamental data analysis provider YCharts, describes how any new software undergoes multiple "rigorous" peer reviews and a "run book" that combines engineers being assigned specific areas of code to test, automated tests, and "user-style tests" that mimic users building a chart or portfolio, before any software enters the vendor's staging environment and is released to an internal test group, and only then to a small group of clients.

The process is well-documented as engineers communicate what works and what doesn't work over communications platform Slack, so that the process is transparent and easily searchable. It's also managed from the top down by CTO Ara Anjargolian and vice president of engineering Kevin Fox, and is evaluated regularly for potential improvement.

"Especially if something slips through the cracks, we'll do a post-mortem, find out where we failed in the testing process, and fix it so it can't happen again," Eplett says.

Sterling Trading Tech follows a similarly rigorous internal process of stress testing—challenging its own order throughput and performance—and regression testing by its dedicated quality assurance team, before making new software available to client service and technical teams, who might spot "real

**Jim Nevotti**
Sterling Trading Tech

**Caleb Eplett**
YCharts

world" client issues that developers may not have thought of. After that, the vendor goes through three stages of releasing software into the world, starting with a limited rollout to early adopters or clients who had requested specific enhancements, then a soft launch, followed by a full production release. Typically, it takes between four and eight weeks for a new software release to go into production, from finishing development to general availability, Nevotti says.

Like YCharts, Sterling prefers to make small, frequent releases, rather than rolling out a bunch of updates all at once, believing this approach ensures greater platform stability because there are fewer variables that could go wrong, while smaller updates are quicker and easier to test.

"It evolved that way naturally, and we've refined it over the years," Nevotti says. The company also evolved toward the agile development methodology, and now tends to work in two-week sprints. "It evolved that way because we have a strong desire to be responsive to clients. When they ask for changes, we don't want them to have to wait six months for the next annual release, when we can schedule it for the next development cycle," he says.

### Clean Up as You Go

John Eley, CEO of data management platform vendor GoldenSource, and senior vice president of product management Tom Stock are also both supporters of the agile model, but have applied several unique twists that they believe give the vendor an edge when it comes to testing.

For example, the vendor has instituted a monthly review of its code quality, where its developers present their progress toward corporate-wide goals to Eley; it surveys clients annually on their perception of its quality; and it has multiple levels of testing during and after the development process, including a group called the "model office," which is independent of its development group and tests software in a way it would be used by clients.

"I don't think we've created any earth-shattering measure of quality, but we've assembled them in a way that's unique to us," Eley says.

The vendor has also made a big investment in its DevOps function over the past couple of years, to make sure any additions don't interfere with existing functions,

Another area of investment over the same period has been on automated testing and tracking tools. Currently, GoldenSource has an employee writing test scenarios, but it envisages being able to automate testing scenarios in the future, which should speed up the testing process, while also allowing its testing staff to focus on more complex scenarios.

"If you can auto-generate some of the functional tests, you can focus more on business testing, and being able to test the way end users would. For example, you may find usability problems that technical test staff didn't know to test for," Stock says. "The 'model office'… is not a generally accepted concept that all software companies follow—it's something we do that we don't believe everyone else does."

### New Approaches

Sometimes, like with Sterling, approaches to testing evolve over time. In other cases, it takes a new hire to bring a different viewpoint and approaches gained from past experience in different organizations.

In YCharts' case, although robust practices were already in place, it was the arrival of Sean Brown as CEO that saw those practices turned into processes. "When Sean came in, he insisted that we created well-documented processes for the things we were already doing," Eplett says.

At Fitch Solutions, the software arm of ratings provider Fitch Group, that person is Derek Ferguson, who joined the vendor one year ago as head of technology in Fitch's IT development solutions division. Prior to Fitch, Ferguson spent 11-and-a-half years at JP Morgan Chase, where he was most recently head of engineering for its commercial banking division, and had previously served as lead order management system developer for the bank's private client workstation.

Over the past year, under Ferguson's tenure, the vendor reengineered FitchConnect—its flagship subscription-based web app for accessing market

research—from being a single piece of software to a set of micro-frontends—the client-facing equivalent of back-end microservices—that coexist within a browser. This has resulted in major improvements to the efficiency of Fitch's development team, and also for customers awaiting the rollout of new features.

Previously, FitchConnect needed to be taken completely offline to roll out new software releases, which could only be done over evenings and weekends. By breaking down the service into a series of micro-frontends, specific functions can be upgraded intraday without impacting other components of the system, and the vendor need only test each new feature being introduced, rather than re-testing the entire system.

"When I came in, our core FitchConnect web application was a single piece of code, and to change one thing, you had to change all of it—and that was the main thing slowing us down," Ferguson says. "For example, every time we released a new feature, we had to re-test everything. And although the testing scripts were good, there are about 20,000 scripts, so it would take two days to run."

His first task was to increase the use of automated testing scripts. "When I arrived, most of the work around testing was to automate processes. Some it was already automated and was part of the build process, and some of it was manual. So to be an agile squad, automation needed to be part of the build process," Ferguson says. Now, using micro-frontends, it only takes a developer five minutes to test an individual new component, allowing them to ship new features much faster, and resulting in a 25% increase in productivity.

Fitch also utilizes two key tools to monitor productivity—BlueOptima, which monitors how much code its developers have added each month, and JIRA, a system that tracks workloads and allows the vendor to review metrics around the productivity of its sprints. "BlueOptima can tell good code from bad code on a technical basis. JIRA tells us something equally important—we may be going fast and building good code, but is it the right code for the business?" Ferguson says.

Automation is also key to Pegasus' plans—not only to instill trust in the quality of its products among potential

clients, but also future plans to make its services more widely available. Pegasus' attention to detail over its testing—it spends equal time writing and testing code—is because it doesn't just plan to sell software, but also to make its source code available to clients, along with its testing capabilities, says co-founder and chief product officer Brian Stephens, who previously served as head of market access, market data, and middleware technology at Royal Bank of Scotland.

"The testing discipline stems from a deep understanding of what clients require after my long experience on the client side. When we built the first version of our APIs in Java, we spent the first few months writing tests, from unit tests that look at one component to integration tests that test certain systems and scenarios," Stephens says. "We've ended up building upwards of 500 tests in each language for our APIs—and that number continues to grow."

### If It Ain't Broke, Break It!

Without structure and automation where possible, testing can descend into chaos. But for some, chaos theory is actually a desirable testing methodology.

**John Eley**
GoldenSource

"
> "If you look at companies like Netflix, Facebook, Microsoft, and others, they are probably using or starting to use Chaos Testing. But in the financial industry, firms are probably only just getting started using this." Derek Ferguson, Fitch Solutions

"Chaos testing in software engineering is the equivalent of crash tests in the automotive industry. It's a very controlled way of looking at a system as a whole … and being able to experiment with a system to increase confidence," says Mikolaj Pawlikowski, software engineering project lead and chaos engineering expert at Bloomberg. "We started doing this in early 2016 because we were working on a Kubernetes microservices platform to allow developers to deploy code quickly. At first, Kubernetes patches were flowing faster than we could deliver them … so we started writing scripts to simulate failures preemptively to gain confidence in our Kubernetes setup."

However, this "chaos" isn't really chaos: it's a carefully organized structure in which one component is something of a surprise attack that engineers must be sufficiently organized for firms to benefit from. "We're trying to push a more scientific approach with more focus on tools and analysis, and less randomness. I think the industry as a whole is moving to that. But it does require a certain level of maturity … and you have to understand what something is expected to withstand before you try to break it."

Fitch's Ferguson is also a proponent of a tools-based approach. In addition to the other tools it uses to support its development and testing, Fitch Solutions has licensed a tool called Gremlin, an enterprise version of the "Chaos Monkey" developed by streaming video service Netflix.

"If you look at companies like Netflix, Facebook, Microsoft, and others, they are probably using or starting to use Chaos Testing. But in the financial industry, firms are probably only just getting started using this," Ferguson says, acknowledging

**Derek Ferguson**
Fitch Solutions

that Fitch itself is among these firms not yet ready to fully unleash chaos.

Currently, the vendor simulates a complete outage quarterly. Chaos testing—when configured with access to a company's AWS account—will deliberately switch off parts of services hosted in AWS or reconfigure services to simulate the impact on network quality.

"Everything we have runs in the cloud," Ferguson says. "If we took that to the next step, we would build it in a way that would be resilient to any outages. We're not there yet, and chaos testing is the best way to do that. But we aren't ready to introduce chaos testing right now as we don't believe the results would be good. … That won't happen in 2021—maybe the year after."

One advantage of chaos testing is that it is the antithesis of an efficient development organization and the way engineers think about code: Code is logical and structured, and their approaches to testing are usually similarly logical and structured, so while they work well for testing expected issues, they don't necessarily account for unpredictable events, whereas chaos is by nature unpredictable.

"You can be preemptive and try to figure out what might break, and test the way you want something to behave—it's not rocket science. But when you introduce randomness, you can spot things you didn't predict," Pawlikowski says.

### Testing Times

Of course, the ultimate random element was Covid-19, which wreaked havoc on firms' and vendors' activities throughout 2020. Pegasus' Stephens warns that with budget restrictions and fewer resources available as a result of the pandemic, new technologies being implemented may not have been tested to the usual levels of rigor demanded by regulators, while not introducing new technologies may result in firms continuing to use legacy systems beyond their sell-by date.

"Specifically during Covid, banks' focus has very much been on keeping the lights on, and a lot of new projects have stopped. That tells me that a lot of these teams are running very lean and don't have the capacity to evaluate and test new services," Stephens says.

**Matthew Storey**
SteelEye

This could impact startup vendors' ability to test code, since some companies rely heavily on clients in their early phases to identify issues in their code while they are still developing functionality. For example, a vendor's approach to testing can change significantly from its startup days to when it reaches a size and scale where it has the resources to establish a full testing group, says Matthew Storey, co-founder and chief product officer at regulatory and compliance technology provider SteelEye.

"In our early days, clients were flagging issues to us … and some clients still like to see early versions of the software," Storey says, which not only provides extra sets of eyes on the software, but also brings the experience and understanding of client practitioners. "Our developers and engineers … might understand a bug in the code, but might not necessarily understand the challenge that the client is facing. … It's important for us all to know what effect it has on a client if we release bad code."

Sterling's Nevotti also highlights how practices continuously evolve as a company matures and changes. "As you grow larger and gain more critical mass, you have more to spend on testing, and also you become more important and critical to clients, so you are held to higher standards compared to when you were a startup," which requires more defined procedures and better testing tools, he says. "A couple of years ago, our testing team couldn't test everything because they didn't have the right tools, so we had to go out and invest in new technologies."

Though SteelEye and Sterling are past that phase, other startups may find the current climate particularly challenging or costly—not just in terms of making sales during Covid, but also ensuring their code is as clean as it can be prior to entering clients' production environments.

One thing's for sure: The more time you invest in testing, the less time your systems are likely to be offline as a result of any errors in their code or issues with your infrastructure. Or, as Bloomberg's Pawlikowski says, "The more you test during the day, the better you'll sleep at night." w‡
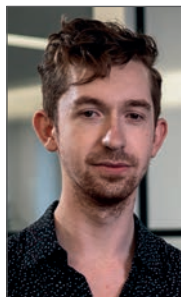
# waterstechnology

# Sign up to *WatersTechnology* newsletters

- Tailor your newsletter experience to suit your areas of interest

- Receive exclusive content handpicked by our editors

- Update your newsletter preferences at any time

Technology

Trading Tools

Operations

Data Management

Regulation

Management & Strategy

Visit www.waterstechnology.com/newsletters to update your newsletter preferences today

# At the 'Fringes of Realism': Agent-Based Models Take Hold Among Quants



Agent-based modeling has taken root seemingly everywhere throughout the last decade, from theoretical physics, to military operations, to public health, to ride-sharing apps like Uber, and to a much lesser extent, finance. However, a year such as 2020—and all its ups and downs—could drive firms to the edge of this new frontier. By Rebecca Natale

Many things changed on September 11, 2001, and apart from the profound sense of loss, there was also stunned disbelief—a question of "How?"

In the aftermath, Justin Lyon enrolled in a graduate program at MIT, studying system dynamics, or, to put it very simply, how complex things—systems with many moving parts—act and change over time. There, he focused on simulating the growth of radical Islam—that is, predicting likelihoods of violence, insurgencies, and counterinsurgencies in tandem with potential consequences that could follow—in Afghanistan, using a technique called complex adaptive systems modeling. Operating under the theory that the universe is composed of highly complicated, in-flux, and interconnected systems, Lyon set out to determine how to model humans and the ways in which they interact in a way that acknowledges and adjusts for all the nuance that colors their rational and irrational behaviors.

Over the better part of the next two decades, Lyon worked as a contractor for a number of organizations, including the Bank of England, ExxonMobil, Microsoft, and the US Department of Defense, for which he ran a team of analysts supporting a general based in Iraq.

Five years after 9/11, but eight years before Lyon would start and head up his own company called Simudyne, the 2008 financial crisis rocked the global financial system, and marked a clear turning point in the prevailing views economists took toward markets—that they were inherently efficient and rational. They aren't, and increasingly, it's dangerous to believe so.

"All models are ultimately wrong because they aren't the real world. But you want to have an ecosystem—a zoo of models," says Lyon, echoing a term

coined by Andrew Haldane of the Bank of England. "A zoo of models: a whole bunch of different, diverse species of models so that you're not blindsided."

When Lyon founded Simudyne, an enterprise simulation technology company, in 2016, his thesis was that people could make radically better decisions that ultimately lead to a safer world. People could do this, in part, by using agent-based modeling, which draws upon his earlier work modeling complex adaptive systems.

The range of possible financial applications for this emerging quant technique is vast. A small universe of whitepapers and research studies pertaining to agent-based modeling and finance has exploded in the last five years, including from Refinitiv, the Bank of England, the Securities and Exchange Commission,

and Deloitte. They put forth cases for its use in areas such as high-frequency trading, execution, market-making, risk management, and predicting and understanding black swans—which, as the future begins to look less and less like the past, may fan the modeling method's slow burn into a blaze.

## Who, What, and Why

As a rudimentary definition, agent-based modeling is a bottom-up approach to simulating the real-world behavior of agents—a single trader on a small scale, the global financial system on a larger one—without basic assumptions that the agents in their environment behave in perfectly rational ways or have perfect information on which to base decisions.

An agent-based-model can be broken down into two components: one that is structural and one that is behavioral, says Krishnen Vytelingum, head of quantitative modeling at Simudyne. The structural piece is the environment where buyers and sellers interact, such as on an exchange, or an actual physical space, such as a city map within the ride-hailing app Uber, where riders and drivers meet. The structure constitutes the rules of the game—the mechanisms players use to connect activities and interact—but the behavioral component accounts for what players could actually do, which can be anything from very simple, even random behavior, to extremely complicated behavior.

With traditional modeling methods used in finance—primarily stochastic factor modeling, which includes Monte Carlo simulations—you model what you observe. And a major shortcoming of traditional methods, says Vytelingum, is that they assure, dangerously, that the failure of one institution is independent from another. Because institutions are inherently connected and exposed to one another, the downfall of one behemoth likely dominoes to another, which can make it extremely difficult to model an accurate view of credit risk.

Stochastic simulations can be quite sophisticated from a mathematical standpoint, but are ultimately glorified endeavors in extrapolation, says Jean-Philippe Bouchaud, chairman and chief scientist of Paris-based quant fund

> " "Traffic jams are really special. You run into a traffic jam, and you don't know why the traffic jam is there, when suddenly it just kind of vanishes into thin air."
> **Jean-Philippe Bouchaud, CFM**

Capital Fund Management (CFM) and a physicist by trade. Instead, by approaching the world through a physics lens, you try to model the behaviors that lead to outcomes.

"Traffic jams are really special," he says. "You run into a traffic jam, and you don't know why the traffic jam is there, when suddenly it just kind of vanishes into thin air."

Given an open road, logical behavior should dictate that drivers will accelerate to the maximum allowed speed and continue that way until they exit the road or until something blocks the path, at which point they would brake to avoid a collision. In a perfect world, the driver would always stop, and cars would never crash.

But in the real world, says Bouchaud, maybe drivers have slightly different response times. Maybe some of them are distracted. Maybe some are prone to admiring landscapes or birds, or are prone to daydreaming. Maybe all of these things are true at once. It may only take a minute or two of lost concentration for a car's velocity to drop enough that a procession of angry commuters forms.

"All of [these little events] can cascade into something major," he says.

Of course, this is a simple example, but it's translatable to the languages of financial crashes, economic crises, or a group of fireflies flashing in unison. All of these are examples of emergent phenomena—that is, collective behaviors—that are difficult to predict.

Ten years ago, CFM, which has more than $7 billion in assets under management, began using agent-based modeling to understand the firm's own impact on the market when it placed market or limit orders. The firm created a fictitious market with market-makers and other types of traders interacting with a



**Jean-Philippe Bouchaud**
CFM



**Lisa Borland**
Cerebellum Capital



**Justin Lyon**
Simudyne

simulated version of the real firm, testing various strategies and their effects. The exercise was useful, Bouchaud says, because the leeway one has to conduct experiments using people's real money is not exactly that much.

Bouchaud was most recently involved in what he calls an exciting new development for agent-based modeling in finance, using it to analyze how liquidity crises are generated sometimes spontaneously and seemingly for no reason, much like the idea of a traffic jam.

Lisa Borland, a senior machine-learning scientist at San Francisco-based hedge fund Cerebellum Capital, joined the financial world after receiving formal training as a theoretical physicist, trying to understand how macroscopic phenomena—think, for example, perfectly structured cloud formations—emerge. Finance offered her similarly complex problems to solve, and she became especially interested in studying the price formation process as one of those complex systems.

For instance, you could create a "toy system" containing a fictitious stock, for which there is a fixed supply. The agents trading the stock could be programmed with simple trading rules—some would trade on recent trends, others on the stock's time series, and others based on the fundamental value of the stock. Then, you would let the simulation run. What you'd find is that, even though you don't have a mathematical way of deriving it, the price would evolve and replicate in the same manner seen in real markets.

"They become great tools if you want to understand how changing the rules would affect the system," Borland says.

Simudyne's Vytelingum found something similar: Even when agents were of "zero intelligence," meaning they bought and sold at random and didn't abide by any trading rules, they produced an efficient market. One could take this to mean that traders don't need to be extra smart or clever, or even efficient, to mimic the kind of rationality typically seen in the market.

But this finding also lends credence to a main criticism of agent-based models—that they aren't calibrated to the real world because the models could be constructed in a multitude of ways

that could each, potentially, give rise to the same phenomena.

## Predicting the Past to Predict the Future

As the 2008 financial crisis was taking hold in the US, theoretical physicist Mark Buchanan penned an opinion piece for *The New York Times* titled "This Economy Does Not Compute," in which he described agent-based computer simulation as a "telescope of the mind" that "[multiplies] human powers of analysis and insight just as a telescope does our powers of vision."

Almost an opposite approach from "what-if" testing—wherein the user identifies a possible event and, based on largely static conditions, looks for how it would play out—an agent-based model gives the user a range of its own possible scenarios, many of which the user likely hadn't ever considered.

On one hand, it could push the boundaries of the ways in which organizations manage and mitigate risk. On the other, because these models are rooted in unpredictability, and the number of possible outcomes vast, one could argue these simulations are mostly arbitrary. In theory, agent-based modeling could teach firms to be prepared for

**Krishnen Vytelingum**
Simudyne

anything, but in practice, they could certainly not be prepared for everything.

However, firms may have little choice but to try because increasingly, the future does not look so much like the past, says Jochen Leidner, Refinitiv's director of research.

Last summer, Refinitiv partnered with Simudyne on a project to predict price movements using a combination of agent-based modeling, synthetic market data, and machine learning. The team, which included Leidner, Vytelingum and two others, worked on the premise that using the past—in this case, historical market data patterns— was a deeply flawed way to predict future price movements.

> "That means we can now no longer just rely on one history—the one that actually happened, the only thing that we really know for sure—but we can also explore the fringes of realism. And we can see how our models behave on these fringes and these fictionalized versions of the almost-past that could have happened but didn't."
> **Jochen Leidner, Refinitiv**

"Whie it's entirely the right thing—so I'm saying nothing against back-testing at all—it is not alone sufficient to deal with the complexity of this world," Leidner says. "You may also want to explore alternative futures that may not be based on the actual history, but on potential alternative histories—so not on the past as it happened, but on the past as it could have happened."

This is a weird notion, he admits. After all, everyone agreed there's only one past. So the idea was to create synthetic data that looked as if it could have been a possible past, in the sense that it shared certain properties that financial markets are known to exhibit, such as seasonality and regularity; these are otherwise known as stylized facts. Real historical market data notwithstanding, the team created purely random—made-up—data that, despite its randomness, structurally and distributionally obeyed the laws of finance as they know them, making it plausible.
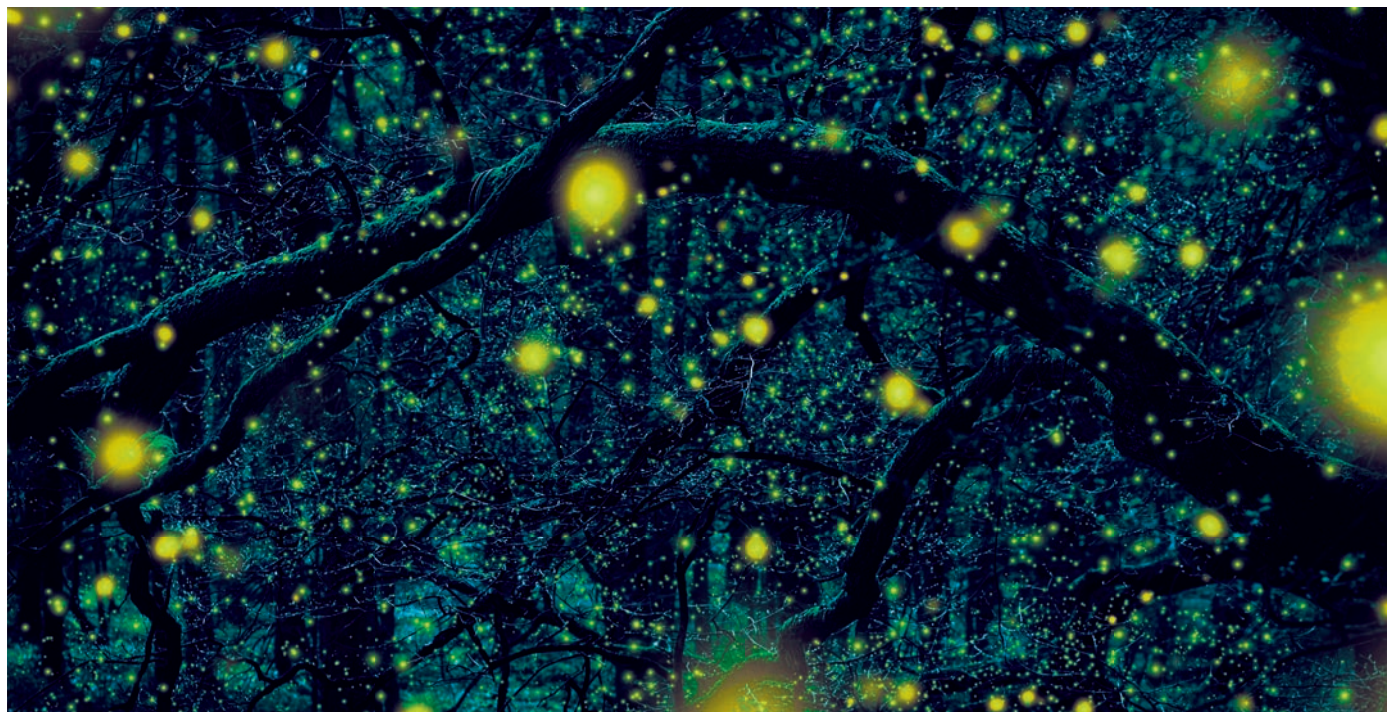
"That means we can now no longer just rely on one history—the one that actually happened, the only thing that we really know for sure—but we can also explore the fringes of realism," Leidner says. "And we can see how

our models behave on these fringes and these fictionalized versions of the almost-past that could have happened but didn't."

While it may seem like this story about agent-based modeling is veering sharply into science-fiction territory, the concept Leidner describes has a very practical application: stress-testing your models in a world that's changing at a faster pace than ever before.

Pandemics are sickening populations at an alarming rate. In the 21st century alone, outbreaks have included SARS, swine flu, MERS, Ebola, Zika, Dengue fever, and Covid-19, in addition to other epidemics with smaller reaches, and experts have warned that global viruses will become more common. At the same time, climate change has become a real, measurable threat with far-reaching implications and an ever-shortening time-frame to counteract it. And more recently, conventional forecasts failed to predict political upsets such as the 2016 election of President Donald Trump, and the UK's decision to leave the EU.

"The world is increasingly running off-script, or deviating from just being a linear consequence of the last couple of years. … That is something that makes it hard to model [the world] just based on the past. And it could therefore be, and that's our contention, that we may want to use synthetic data in a wisely-applied form to expand our toolkit for stress-testing our models that we may want to create for trading or compliance," Leidner says.

## War Games

If, at the least, agent-based models aren't yet useful at the event-prediction level, they can serve as a kind of war game, or tactical exercise, for banks and traditional asset managers that don't yet utilize them. Though today the models are concentrated among a handful of quant funds, Barclays, which led a series A round of funding in Simudyne, is one example of an early adopter. C.S. Venkatakrishnan, Barclays' global head of markets, has written about the bank's projects with Simudyne's software, but was unable to comment in time for publication.

CFM's Bouchaud, who has been working in this field for the last 20 years, says physicists and financial quants are indeed already thinking up the next evolution of models, which he says would be to have an agent-based model that's accurate enough to tell a trader whether they should buy Apple or PepsiCo, for examples.

"That would be great. But before we get there, I think that the mere fact of having these scenario generators is extremely important from a risk-control point of view," Bouchaud says. "It is hard to guess the right direction in which stocks are going to go, but if you can avoid major events, that's already something good." wt

# NLG: The Next Big Thing in Data Analytics?
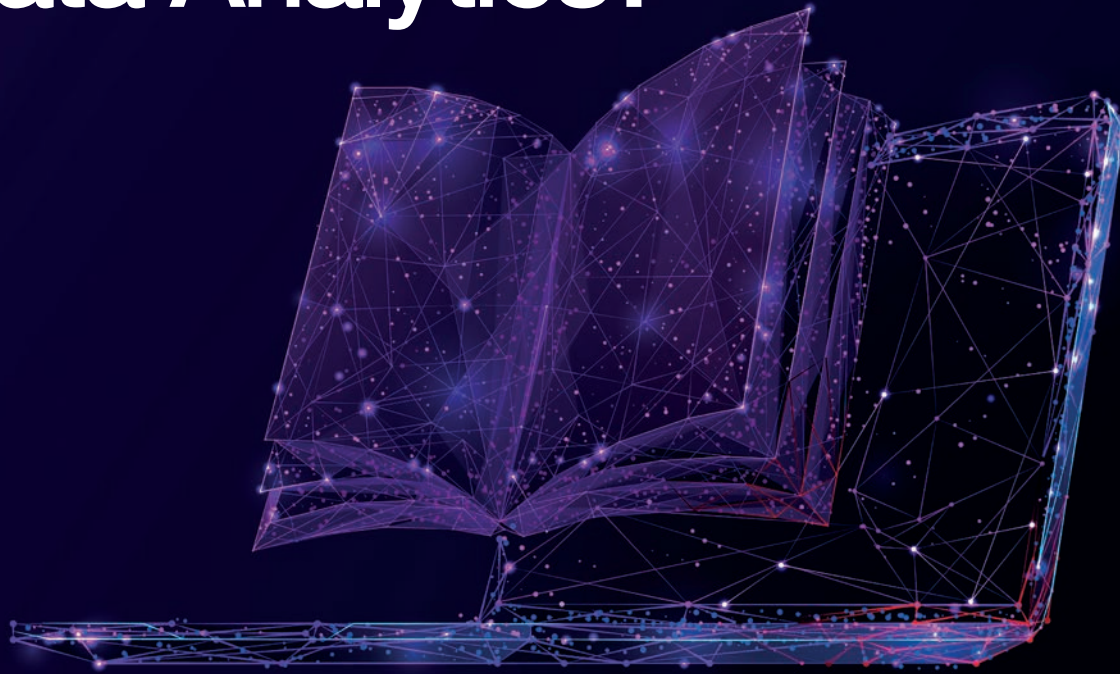
Sell-side firms and data providers are increasingly experimenting with natural-language generation to create new forms of automatically curated reports, emails and alerts. As natural language processing becomes more widely adopted in the capital markets, NLG should follow suit—but the technique comes with significant challenges. By Josephine Gallagher

If only data could talk—well actually, it can.

Over the last decade, thanks to the proliferation of open-source tools, improved computing power, further cloud adoption, and major advancements in the retail space, natural-language processing (NLP) is becoming democratized, and capital markets firms are taking notice. The next evolution of NLP, though, is natural-language generation (NLG). While the technology has been around for decades, it is only in the last two years that it has begun to make meaningful inroads in financial services.

During a fireside chat at this year's WatersTechnology Innovation Exchange, Kim Prado was asked to name one trend that will bring the most value to her business in the next five years. The global head of client, banking, and digital technology for Royal Bank of Canada's Capital Markets group, responded, "Definitely taking natural-language generation to the next level."

While there is a lot of excitement around this improving technology, as Prado notes, it needs to get to "the next level" if it is going to live up to its promise and help traders and portfolio managers find valuable information on which to execute.

Those singing NLG's praises say it will become a game changer for those trying to make sense of copious amounts of data flashing across their screens, but others are more reserved in their predictions of what it can achieve. Tim Nugent, senior research scientist at Refinitiv, who specializes in machine learning, natural-language processing, blockchain, and cryptocurrency research, says that through internal research and testing, Refinitiv found that extractive techniques—which pick out the most salient words or phrases directly from the original text to create a summary—are more effective than NLG at summarizing text.

"By coming up with an NLP-based scoring function at the sentence level, we think we can create better summaries, [rather than] by applying some type of NLG approach," Nugent tells *Waters Technology*. "And I think when

we benchmark our methods and score them, at the moment, it's obvious to us that extractive approaches are superior."

Adding to the attention around NLG are the significant breakthroughs that have been achieved in the last couple of years—most notably with the advancements in pre-trained language models such as the latest release of the release of Generative Pre-trained Transformer 3 in June (*see box, page 25*). Paul Tepper, an executive director in Morgan Stanley's technology division, says that while these advancements are impressive, heavily regulated financial institutions need to be mindful of the risks associated with advanced pre-trained language models.

"I would be pretty shocked if any large company let one of these things loose on their customers, or in any other specific case, because you really don't know what it's going to do, like it is a probabilistic model. And it is basically just generating the next word based on things it has seen already, so it is quite unpredictable. We are not using anything like that," Tepper says.

### 'Low-Hanging Fruit'

From 2008 to 2013, Vicky Sanders was the global and European head of equity sales at Goldman Sachs. Back then, she arrived in the office at around 6 am, spending the bulk of her mornings drafting summaries based on the latest events impacting the equities portfolios she covered. Those summaries were sent out to client analysts' and portfolio managers' inboxes ahead of their workday. Today, Sanders sits on the other side of the tracks as global head of investment analytics at Liquidnet, where her team is using NLG to automate this cumbersome task.

Liquidnet uses NLG to automatically create customized email alerts for a specific portfolio, which are then sent to portfolio managers and traders throughout the day. The NLG capability sits within Liquidnet's Investment Analytics suite, which was created after the acquisitions of RSRCHXchange (of which Sanders was co-CEO), Prattle, and buy-side analytics platform provider OTAS Technologies. Leveraging portfolio managers' watchlist, Liquidnet's system can alert users to relevant research reports from across its

research library, which includes more than 430 research providers. NLP is initially used to comb through a further library of content including bank records and public information, such as earning calls and press releases. Once the most important information is extracted, the NLG then kicks in and converts the data into email alerts written in natural language.

Sanders says it is these types of repetitive, low-skilled, and time-consuming tasks that are ripe for NLG.

"There is still a huge amount of low-hanging fruit in our industry to use things like NLG to solve for—what can only otherwise be described as manual labor," she says. She adds that NLG is best placed to handle simple tasks that will help free up teams on both sides of the Street to do more with less.

Tom Doris, chief data scientist at Liquidnet—who was formerly CEO of OTAS—points to another example of how NLG is automating tasks. Historically, traders or portfolio managers would have a quant sitting by their desk crunching numbers to surface actionable insights. He says "the human in the loop"—meaning the quant, in this case—was there to simplify quantitative analysis and make it easier for the front office to understand. Today, many traders and portfolio managers now have automated analytics charts and alert systems lighting up their screens.

However, Doris says that today, NLG is helping to turn those colorful analytics charts into automatically generated, readable information. For example, if the spread or liquidity profile of a stock changed significantly intraday or a company's stock price drastically moved compared to its competitors, he says these types of market shifts can be summarized in one or two sentences—meaning there is less "subjectivity" or "mental workload" for the trader to cope with.

"Our fundamental finding was that it's easier for us to do the work to convert the content to NLG text than it is for every single one of our users to interpret charts, tables of data, and numbers in the heat of battle, so to speak. So, it is just simply a more efficient way of communicating content to humans," Doris says.

> "Our fundamental finding was that it's easier for us to do the work to convert the content to NLG text than it is for every single one of our users to interpret charts, tables of data, and numbers in the heat of battle, so to speak. So, it is just simply a more efficient way of communicating content to humans." **Tom Doris, Liquidnet**

**Tom Doris**
Liquidnet

**Vicky Sanders**
Liquidnet

Liquidnet's NLG technology is rules-based and includes a finite list of sentences and paragraph templates for the system to choose from when converting data into natural language. While the technology model can be trained to turn structured data and routine events (i.e. price shifts) into bite-sized summaries, Sanders says more complex readings of data should be left up to the human.

"When it comes to interpreting a [complex] chart, that's probably where you'd want to lean more on the human intelligence and the artificial intelligence (AI). … I think for our industry and the use cases we're looking at, most of it is still in the low-hanging fruit area, as opposed to further up the value curve or intelligence curve," she says.

### A Character Issue

Ted Merz, global head of news product at Bloomberg, describes NLG as a proxy—in other words, a single cog in a complex engine that turns data into human language. He says the data and media giant's readership of NLG-generated news articles, from multiple media outlets and accessed through its Terminal, has gone from zero to 7%, in the last two to three years.

Bloomberg is using NLG to produce four types of automated news articles: stories compiled from alternative data sources, corporate filings (i.e., earnings announcements), combined datasets (a mixture of traditional and alternative data), and market anomalies.

Taking the first group of articles, these are generated from "obscure" alt

datasets, as Merz calls them, such as the number of people riding the subway, the number of people working in an office, or the number of restaurant bookings via applications such as OpenTable. Merz says portfolio managers are less familiar with these types of datasets, so turning them into articles makes them easier to understand.

"If you were only looking at the dataset from OpenTable about restaurant bookings, it would be hard to understand what was happening. So, what we do is write scripts [or articles] that say, 'The number of restaurant bookings are up or down versus comparable periods; these are the areas and places where they're increasing or decreasing the most,' and it really makes that data much faster and easier to absorb," he adds.

Like most NLG systems, Bloomberg's is based on a set of rules. Within that, it also has thresholds for when an automated article should be published. For instance, take the third group of news stories, generated from combined datasets. In the event of a hurricane, the tool would pull in datasets such as weather information and geolocation data to predict what company assets could be at risk of being impacted. As an example, Merz

**Ted Merz**
Bloomberg

says if the hurricane were to hit 25% of Exxon's oil facilities, the risk percentage would be enough to trigger the system to publish the article. The logic behind this is to avoid spamming traders or portfolio managers with meaningless information.

When building an NLG system for summarization, one of the hardest things to get right is data quality. Merz says when building its summarization tool—which summarizes multiple news articles—a big decision the development team had to wrestle with was the maximum length for the summary. Although this sounds like a trivial challenge, he says the length of the characters can drastically change the way the NLG generates a summary.

The team found that if the length was too short, it did not make sense; if it was too long, it was no longer that useful as a summary.

"If you said the length was 10 characters, 30 characters, or 60 characters, you would get a different summarization every time. It is not like you just add characters; the computer recalculates the summary completely differently," he says.

## A Matter of Understanding

Tepper has more than 20 years' experience working with NLG and has

written his undergraduate and master's theses on the technology. Today, he focuses on AI and NLP for wealth management applications at the bank.

While most firms exploring NLG have been using it to summarize data, Morgan Stanley has taken a different route by embedding it in its chatbot for assisting its financial advisors. The bank has an internal contact center that fields calls to its financial advisors and their support staff, but the hope is that with the development of a chatbot assistant, the NLG feature could alleviate some of the workload before reaching the advisor. The advisory chatbot is separate from Morgan Stanley's AskResearch bot that was built to help bank analysts and sales teams query thousands of reports generated each year.

"The call center internally fields, like, millions of calls a year—not billions—but millions of calls a year. So, it's a significant cost and if we can divert some of the cost by answering the question with this self-service system, or we can sort of reduce the amount of time they spend looking for these answers by getting them part of the way there without having to

talk to a person, we can both reduce the cost of running these system, as well as potentially provide a better experience for our users," Tepper says.

Tepper says the bank is about six months into building the advisory chatbot and has trained it with several hundred intents—meaning the bot should be equipped to answer hundreds of questions. The bank also has plans to extend the NLG function to client-facing applications like Morgan Stanley Online, its web-based and mobile-app interface.

Tepper says the most challenging part of building NLG-based tools—or even NLP, for that matter—is developing the ontology, otherwise known as a knowledge graph. The ontology organizes all of the information that the chatbot would leverage to understand the intent of a user's query—including documents, legal entities, types of accounts, or types of businesses. Once the machine understands the user's intent, it can prompt a dialogue.

To illustrate this further, Tepper says if an analyst typed "individual retirement account" into its chatbot, the chatbot would be prompted to ask a variety of questions, such as, 'Do you want to open an IRA account?', or, 'Do you want to close an IRA account?', and so on. Highly skilled knowledge graph engineers—who are in significant demand these days from Big Tech and financial firms—are responsible for building these complex ontologies and transferring domain expertize into knowledge representations.

"The challenges are getting that knowledge out of people's heads, making sure it's correct and precise, and that the language being generated sounds natural," says Tepper.

Tepper says Morgan Stanley's NLP and its subset natural language understanding (NLU) applications—which focuses solely on the machine's ability to read and comprehend—incorporate machine-learning techniques. Yet the NLG is based on traditional symbolic rules—and for good reason, he says.

"The trouble with the NLG being machine learning-driven is that if it's actually generating what to say, then you're going to be trading off some of the control—you won't know necessarily what it's going to say. So you don't necessarily want to be in a situation where you're generating content that somebody is going to read and then go forward with, [but] you don't know what it is potentially going to put together [and] what it's going to say. So, you have to be kind of restricted on how you can use machine learning there," Tepper says.

> ❝
> **"The challenges are getting that knowledge out of people's heads, making sure it's correct and precise, and that the language being generated sounds natural."** Paul Tepper, Morgan Stanley



**Tim Nugent**
Refinitiv

### Abstractive vs Extractive

Refinitiv's Tim Nugent splits data summarization into two categories: abstractive and extractive. NLG falls into the abstractive category, meaning it generates an abstract summary based on a fixed language model—or a finite number of sentence structures—and the data it is tasked with summarizing is used to fill in the gaps. Take this sentence for example: "The S&P rises by 1% today." The words "rises" and "1%" could change depending on whether S&P rose or dipped that day; however, the rest of the sentence could be fixed. This is a basic example of how NLG generates a sentence.

In the extractive group, NLP is used to pull lines directly from an article to form a summary. Nugent says Refinitiv has experienced more success in using this approach than NLG.

"Without any fancy bells or whistles, a very strong baseline approach to summarize an article is to actually take the first three sentences, and this is because journalistic style tends to dictate that you put the most important aspects of the entire article up front," he says.

His view is that summaries produced by NLG are more likely to omit important bits of information in the article or document than using an extractive approach. He says factors like summary length and the length of the inputted documents are contributing factors in overwhelming the NLG, and which effects the quality of its output.

"If you were to drop the entire earnings call transcript into an NLG approach, you've got so much information, you've got so much text from which to source your abstractive summary from, you are much less likely to capture the critical sentences than, for example, if you apply this extractive approach. So the length

## GPT-3: Breakthrough or Calamity?

One of the recent developments in the world of NLP and NLG that has brought about a mixture of excitement, buzz, and concern is Generative Pre-trained Transformer 3, or GPT-3.

The autoregressive language model uses deep learning to produce natural language text. It was created by OpenAI, a research business co-founded by Elon Musk and is pre-trained on a vast corpus of data—about 175 billion parameters.

While the technology can be applied to multiple tasks—such as generating summaries or even writing code—some are skeptical of applying it to real-world applications, particularly in the heavily regulated financial services industry.

"We just absolutely refuse to even run that risk," says Tom Doris of Liquidnet, in talking about GPT-3 and whether the firm uses it in its own NLG summarization tools. Rather, the firm's NLG is built on a rules-based engine and proprietary ontology. "The solution that we've come up with works very well for us, without needing to train on huge corpuses of un-curated content."

Tim Nugent of Refinitiv is also a self-proclaimed skeptic of GPT-3. He agrees with the view that it still is not well understood how the technology works. To illustrate this, he uses an example of a GPT-3 model being tasked with summarizing an article in 50 words. He says thatj although the model may use the article to inspire the summary, it also draws on an unknown corpus of pre-training data, thus making it difficult to predict what information it could spit out.

"You actually have much less control over what GPT-3 generates than you might think, and that absence of control should absolutely be concerning for customer-facing output," Nugent adds.

Nugent believes that there are fewer safeguards when using GPT-3 and protecting clients from noisy data, potentially inappropriate language, and just data of a poor quality.

Paul Tepper of Morgan Stanley says pre-trained neural networks that underpin GPT-3 are impressive—but echoes similar concerns about exposing clients directly to the technology. He says he is doubtful that any major organization would have "the appetite for unleashing unsupervised learning on its customers."

To drive home this point, he says institutions are heavily regulated entities and are limited as to what they can do with machine learning or AI. There are global guidelines, such as the European Parliament's guidelines on ethics, which require financial services firms to explain how their algorithms and AI work.

Kim Prado of RBC says the bank has not yet explored GPT-3 in its NLG. "We have not tried out GPT-3, but we are actively watching the adoption in other areas. GPT-3 is closed source and we are yet to onboard it," she says.

> **"Our focus was on reducing the data 'firehose' effect, and providing key actionable insights from data, and in this case, NLG was able to solve for that business problem with an impact."**
> **Kim Prado, RBC**

of the input document actually has a fundamental impact on the quality of the output depending on whether you choose an NLG abstractive approach or an extractive approach," he says.

### Story Time

RBC's Kim Prado, on the other hand, sees NLG as a tool for telling a story about the bank's data. RBC opted to use NLG over other summarization techniques following multiple proofs of concept over the last several years, she says, where the team has become comfortable using the technology.

"Our focus was on reducing the data 'firehose' effect, and providing key actionable insights from data, and



**Kim Prado**
RBC

in this case, NLG was able to solve for that business problem with an impact," she says.

RBC is now building a summarization engine for turning unstructured client-interactions data into human-readable summaries. For the technology to work, the firm's NLP first identifies important client-interactions data—such as product mentions or followup activities—from sources like Salesforce, its client relationship management system, and proprietary and vendor chat applications.

The NLG then converts the interactions data into natural-language summaries. Those summaries are then pushed out to users of the data, in this case the sales team, traders, product owners, and even senior executives.

"We are hoping that by providing a clean and relevant summary, it gives our users the incentive to better use our systems, and eases their minds in needing to browse through pages of reports in order to make sense of what is happening," Prado says.

For building its NLP, the bank has used Google's Bidirectional Encoder Representations from Transformers (BERT), an open-source transformer-based language model, as well as model embedding, sequence-to-sequence transformer models, and classical machine-learning techniques. The bank has no plans to train its NLG using BERT or other advanced encoder-decoder models, says Prado, but the team is open to experimenting with new ideas as they mature.

Regardless of the camp you're in when it comes to NLG—whether you think it will revolutionize dashboards or play a small role in how portfolio managers consume information or how chatbots operate—it is clear that NLG is making inroads in financial services.

The way traders and portfolio managers consume data and the battle over desktop space has been a major focus end-users and vendors alike since the advent of computers on Wall Street. The big test for NLG will be whether the front office finds an appetite for natural-language summaries on their screens, or opts for flashing visuals and NLP-powered analytics. **wt**

# waters tech
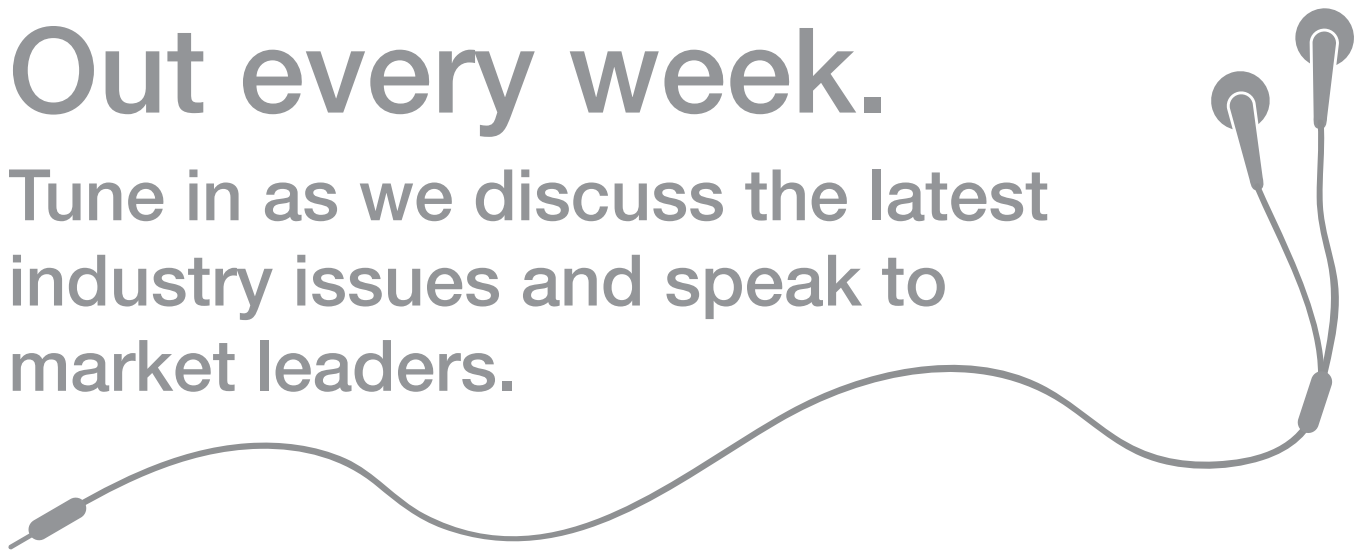
# Waters Wavelength Podcast

## Out every week.

Tune in as we discuss the latest industry issues and speak to market leaders.

# Regulation Won't Address Market Data Costs

Market participants say Esma's latest efforts to address the cost and complexity of market data fees don't address the root of the problem. By Joanna Wright

European guidelines issued in November, the latest regulatory attempt to address what market participants say is the high cost of market data, might help to harmonize some existing practices at trading venues. What they won't do, though, is address the root causes of these high costs, or the complexity of market data usage policies and licenses, market participants and practitioners say.

"This document is a way of trying to satisfy everybody, but really ends up satisfying nobody," says Aquis Exchange CEO Graham Dick.

In November, the European Securities and Markets Authority (Esma) released regulatory guidance in the form of 16 draft guidelines for public consultation, aimed at helping trading venues and market participants understand how regulation applies to keeping market data costs reasonable, and pricing policies transparent. But sources say the guidelines will not have any impact on costs because they do not address the complexity of policies or the fees that trading venues are now levying for so-called "non-display" data and "derived" data.

## Regulatory Flashpoint

Regulators must be used to pleasing no one when it comes to their efforts around market data costs. This is a volatile flashpoint for them—there are just so many vested interests and competing claims involved.

Consumers of market data have a list of complaints: Exchanges have always had effective monopolies and massive market power because of the exclusive data they produce as a by-product of their activities; exchanges want to charge for data based on its worth, rather than on a "reasonable commercial basis," (or, in other words, fees that cover the

cost of producing and disseminating the data, plus a reasonable margin); and the Markets in Financial Instruments Directive (Mifid and Mifid II) and its delegated regulation, Mifir, have failed to bring down high costs.

These consumers complain that as they are increasingly taking in and using market data in new ways—routing it to execution algorithms and risk calculations, deriving their own products from it—the trading venues try to muscle in on what they see as new revenue sources by applying new fees and new policies, which have evolved to be opaque and of Byzantine complexity.

Exchanges, on the other hand, say they have businesses to run and bottom lines to protect. They have high overheads, and must invest in robust, resilient connectivity, data infrastructure, and human resources. And, they say, contrary to market participants' claims, exchange fees have not rocketed up in the past few years. Rather, expenditure on data has increased because companies are buying more: Market fragmentation and automated trading have increased the volume at which market data is needed.

### Guide to the Guidelines

Having canvassed these opinions, Esma put out a consultation almost a year ago that assessed industry views of the cost of data since Mifid II and Mifir came into force. The regulator also said at the time that it would issue guidance for exchanges on the requirements for trading venues that would bring down costs, a promise that resulted in the November guidelines.

The 16 draft guidelines clarify the Mifid II and Mifir rules that exchanges must provide data on a non-discriminatory basis, that they charge for market data on a per-user basis, that they unbundle data for consumers, that they make prices and terms and conditions transparent and easily accessible, and that they provide data free of charge 15 minutes after publication.

Michael Thomas, a partner in the financial services team at law firm Hogan Lovells, says there is quite a variation in the ways that market data can reach end users, and these draft guidelines are trying to introduce a consistency of approach, regardless of how the data is consumed, to ensure that recipients using it for the same kinds of use-cases are treated equally.

"Mifid II sought to introduce a more level playing field for all aspects of the financial services sector, including, importantly, the access to data. So the aim of these guidelines is to ensure that there are consistent standards and approaches for the market, being able to access the market data they need in order to inform their own trading activities and strategies, [and] to inform the design of products that rely on market data," Thomas says.

For example, the guidelines say that customers buy market data to use for multiple purposes like research, index production, or portfolio management, and are sometimes required to pay multiple times for it.

"Where data use-cases are not clearly predefined, such practice renders it very difficult for customers to understand which fees are applicable to them. Furthermore, a basic fee is sometimes added by default regardless of the use made of data, inflating unnecessary [sic] the price," the guideline consultation document says. "Esma is of the view that customers should not be required to pay multiple fees for the same data."

> "These guidelines are trying to set common standards, and to make sure that market data providers are not able to manipulate the basis on which different types of customers can access the data without good justification."
> **Michael Thomas, Hogan Lovells**

Guideline 4, therefore, requires market data providers to consider where customers could belong to more than one customer category and make sure they are classified as one type only.

Or, in another example, guideline 6 clarifies that trading venues must charge on a per-user basis—not per-screen. If a trader uses, say, a Refinitiv Eikon and a Bloomberg Professional service, they should not be charged twice for viewing the same data on both.

"These guidelines are trying to set common standards, and to make sure that market data providers are not able to manipulate the basis on which different types of customers can access the data without good justification," Thomas says.

Most of these draft guidelines are already practiced, more or less, by all trading venues in Europe—what the guidelines are trying to address is the "more or less." Guideline 6, in the example above, is standard practice now at larger exchanges, but there might be one or two smaller ones in smaller member countries that still charge per

**Michael Thomas**
Hogan Lovells

screen, and the draft guidelines make it clear that Esma will enforce the rule to a particular standard.

"Charging per screen [rather than per user] became undefendable about five years ago," says one market data expert who has worked at both large banks and exchanges. "Nearly all of the exchanges have stopped using that and have moved to the per-source, or the 'bums-on-seats' model. And they seem to be more willing to simplify and charge by the number of people using the data, mostly because they are making money out of other policies now."

Janet Mail, head of commercial management at market data consultancy CJC, says guideline 11, which says that market data providers should adopt regulators' standard definitions in their market data policies and price lists, will be helpful. Some of these terms, like "customer," "derived data," and "non-display data," may seem pretty basic, but their definitions are not harmonized across policies; different trading venues have slightly varying framings of what these terms mean.

> "
> **"It doesn't truly address the problem, and it doesn't simplify or make market data more transparent at all."**
> Graham Dick, Aquis Exchange

Euronext, for example, includes derived data under its definition of "original created work"; Deutsche Börse incorporates it in its definition of "information."

These standard definitions are a step in the right direction toward harmonizing the policies of the 40 or 50 different trading venues in the EU, Mail says.

"The standardization of definitions—what derived data is, what an application is—will help," she says. "Right now, a user can't just say, 'Right, [we should] use the derived data,' because they would have to check that against 40 or 50 derived data policies, each of them slightly different. What I consider to be derived data might fall into

**Graham Dick**
Aquis Exchange

the definition of derived data for 15 exchanges, but it might fall outside of the definition for 15 other exchanges. If there was standard wording, that would start the alignment of policies."

### Pleasing No One

But, Mail says, while these guidelines are just a step in the right direction, "they won't have much impact." Even if there are aligned definitions between 40 different policies, there are still 40 different policies.

Dick from Aquis agrees that the draft guidelines are not enough.

"There have been loud cries for transparency in market data pricing because it is considered to be opaque. Every different provider of market data has a different pricing policy, and they can run into 60, 70, 80 pages. So where there is some standardized template for how market data pricing is done, whether that will change anything, I doubt it. And we will probably have a 70- or 80-page template anyway to try and fit in what everybody wants to get in," he says. "It doesn't truly address the problem, and it doesn't simplify or make market data more transparent at all, from what I can see."

He adds: "The issue is that trading venues have implemented so many kinds of licences that it has got into this ridiculous level of detail."

The reality is that the ways in which market data is consumed, and the uses to which it is put, have evolved, and exchanges have evolved their policies along with them. When data used to go principally to terminal screens, exchanges could charge per screen or per user—the "bums-on-seats" model. However, about 10 to 12 years ago, data usage began to shift away from display usage—prices flickering on a screen—more and more to non-display purposes, for feeding into execution algorithms or smart order routing or pre-trade risk calculations. Because no human ever looked at this data on a screen, there were fewer users overall, and fewer bums on seats to charge for.

The Federation of Stock Exchanges (FESE), in a response to Esma's December 2019 consultation, quoted research that demonstrated how dramatic this shift

has been at some firms. Goldman Sachs' cash equities trading floor employed 600 traders at its height in 2000. Today, there are just two equity traders left.

"Complex trading algorithms, some with machine-learning capabilities, first replaced trades where the price of what was being sold was easy to determine on the market, including the stocks traded by the Goldman 600. These new data users (quant, robotic, and artificial intelligence systems) require constant investment in hardware and software by data providers in order to keep up with the new technologies used by these systems," the FESE said.

Trading venues were losing out by failing to capture these new applications of their data. So they started adding non-display policies to their market data pricing policies.

The London Stock Exchange introduced its initial non-display policy in 2010. In 2012, the Oslo Børs began charging separate fees for clients that used its data in non-display applications, in line with a similar policy at Deutsche Börse. In 2015, the New York Stock Exchange (NYSE) scrapped a policy for certain data feeds that allowed users to net their costs and pay once for using the data in multiple display devices, while introducing new policies on non-display usage to better reflect the increase in non-display use of the data.

Somewhat more recently, trading venues have tried to capture revenue from derived data, as users increasingly turn the market data they ingest into their own products, such as indices or pricing models, by running it through their own models, combining it with other data, and otherwise manipulating it.

These new licenses have incensed market data consumers, who say that trading venues have no production costs associated with a market participant's derived data products.

In April, US trade body the Securities Industry and Financial Markets Association (Sifma) vocalized the opinions of market participants on both sides of the Atlantic when it complained about "a new business practice whereby exchanges force market data purchasers to consent to a licensing interest in

derived market data. … Unlike other, typical contract negotiations, market participants are forced to agree to such terms as a condition to purchasing the market data from an exchange that is an exclusive purveyor of that data."

Exchanges respond that they have to make money somehow, and their data is of huge commercial value to the market. As the FESE put it, "Complete business models are based on the valuable reference price data which is provided by exchanges on a non-discriminatory basis. … Exchange data fees represent production costs for third parties' business models, rather than a regulatory requirement, as often presented."

Broadly speaking, EU exchanges will probably support most of Esma's guidelines, though there are some stipulations that could produce additional administrative costs. If they demonstrate compliance with Mifid II's reasonable commercial basis requirement, regulators may choose to ignore calls from market participants for much blunter instruments. Some dealers have called for revenue caps based on the "long-run average incremental cost" model that is sometimes used to regulate the fees of telecommunications providers with significant market power.

## The Real Solution?

For some in the industry, a European consolidated tape—particularly a post-trade tape that is cross-asset, cheap, or even free, and as close to real-time as possible—is the real answer to market data costs, and the complexity of market data contracts. Esma's consultation included questions on why a tape provider failed to emerge in Europe, despite regulatory provisions in Mifid II. But proponents are optimistic that regulators and lawmakers will address the tape next year, perhaps with changes to actual Mifid II legislation.

Jonah Platt, US head of government and regulatory policy at Citadel, says that "even if these [Esma] guidelines were fully implemented, we think that a consolidated tape of post-trade data is just fundamental to allowing market participants to be able to achieve an aggregate view of trading activity across the continent."

However, he says, "There must be mandatory contributions to the consolidated tape provider (CTP) from each trading venue and approved publication arrangement (APA). Just as it's not viable for market participants to go around and negotiate individual data

agreements with each trading venue and APA in the EU, it would not be viable for a CTP either."

Dick from Aquis is also an enthusiastic supporter of the idea that where Esma's regulatory efforts fail, the consolidated tape will succeed.

A consolidated tape of market data would "create a single golden source. This is important because—principally because of all these market data contracts—you still don't know what a particular security has been traded [and] where on a real-time, consolidated basis. The very top professionals can take and source all of that data all of the time, so they know. But everybody else doesn't. So it would simplify the data, but also simplify the contractual distribution of the data," Dick says.

"If you were to implement that, a large percentage of everyday users could sign a single contract with a single entity that would cover all their usage on a pan-European basis. That is very much where I would like Europe to go in the long term," Dick says. "Where we are today is that it's a series of additions and modifications, piling up old policies on top of new policies on top of old policies that have created these monster market data agreements." **wt**

# 'MASSIVE LAND GRAB': Users Fear Price Hikes After S&P Buys IHS Markit

S&P's purchase of IHS Markit reflects the broader trend of market participants pursuing scale to create true front-to-back trading and data environments, which may signal a trading platform acquisition in the future—but the IHS Markit buy may face regulatory hurdles first. By Max Bowie, with additional reporting by Josephine Gallagher, Rebecca Natale, and Anthony Malakian

Index and data provider S&P Global's proposed acquisition of data vendor IHS Markit has garnered praise from analysts and investors, but has left clients nervous that the deal will not see cost savings passed on to end-users, and will ultimately result in price increases.

The deal—announced November 30, and expected to close in the second quarter of next year, subject to investor approval and clearing any regulatory hurdles—will combine S&P's famous equities index benchmarks with IHS Markit's fixed-income indexes and other fixed-income data; S&P's Cusip identifiers with IHS Markit's Reference Entity Database; and the wealth of legacy IHS Markit information assets with S&P's data, analytics, and data distribution mechanisms. S&P also owns energy data provider Platts, while IHS Markit has strengths in environmental, social, and governance (ESG) data. Both energy data and ESG are viewed as high-growth areas

that, along with alternative data and other "adjacencies," offer a $20 billion addressable market that is growing at 10% per year.

"ESG data and insights are increasingly important to the markets and our customers, providing crucial information that investors, risk managers, corporates, and governments need to make decisions every day," said S&P Global CEO Douglas Peterson on a conference call announcing the deal. "This

## 121

Lance Uggla founded Markit in 2003, making 29 acquisitions before merging with IHS in 2016, which had made 86 acquisitions. The combined vendor since made six more acquisitions prior to agreeing its sale to S&P, totaling 121.

transaction will make the company a premier resource for ESG insights and transparent, robust, and comprehensive product solutions."

The acquisition solidifies S&P as the third-largest data provider overall. It also sees it leapfrog Refinitiv and Bloomberg in terms of sales to bank clients, with an estimated $477 million in bank sales, versus Refinitiv with $490 million in sales of data to banks, and Bloomberg with an estimated $419 million in data sales to banks, according to research from Burton-Taylor International Consulting. Burton-Taylor believes the union will "challenge" those market leaders over the next decade and "provide greater market intelligence to offer new trading signals and back-office savings for its customers."

> "There is a massive land grab to become the central platform. It's all about scale and being the platform for full front-to-back trading, from idea generation, to portfolio construction implementation, to trading and execution, to middle-office processing, to back-office regulatory reporting."
> **Brad Bailey, Celent**

However, other observers caution against direct comparisons, noting that the combination "does not present as a logical competitor to Bloomberg and Refinitiv" because it lacks a large-scale real-time data business. Rather, they say, the deal presents a new avenue for vendor growth—gaining massive scale in "niche" areas without competing in low-margin business like equity terminal sales.

"I think the S&P–IHS deal is interesting … [and] is a good fit—a better fit than, say, IHS merging with Bloomberg, Refinitiv, or Intercontinental Exchange," says Mark Hepsworth, CEO of data management software vendor Alveo (formerly Asset Control), who previously ran the Comstock real-time business at S&P, which was sold to Interactive Data

(now part of ICE). "I don't think this is about them really taking on Bloomberg and Refinitiv, as … in many cases data vendors are specialists in particular areas and often dominate those. This is more about what they can do with a huge inventory of data and distribution, and making it easy for clients to consume."

Burton-Taylor also notes that the vendors' index businesses are particularly complementary, and that the IHS Markit indexes, which bring in around $61.7 million annually—twice that of S&P's fixed income indexes, though S&P makes around $1 billion per year from its index business overall—will receive "a significant boost from S&P's brand, sales distribution, and connections."

Indeed, S&P has identified $350 million in revenue synergies, which

it expects to achieve through new products and cross-selling. The vendor has also identified around $480 million in cost synergies—with business overlap and efficiency accounting for 65% of that target, corporate functions and technology accounting for 25%, and real estate accounting for 10%. Sources say those cost synergy numbers are high. However, nervous employees may be comforted by the lack of overlap, and that Markit already operates a very "lean" model, so any "bloodletting" may be limited.

### Pricing Power

Even before the deal was confirmed, Deutsche Bank's analysis of the acquisition identified "significant" cost and revenue synergies, and gave S&P a "Buy" rating, with a target price of $415 (compared to $341.08 on Friday, November 27, prior to the deal being announced, and $326.61 at market close on Thursday, December 3). But while banks' analysts were lauding the deal, their colleagues in market data departments were less optimistic.

"M&A in this industry is rarely beneficial for customers," says one market data manager at a European asset manager. "The most welcome ones are … where there is not much crossover, such as a data company buying a technology provider. But when one fish swallows another, it just creates more dominance and less end value for clients."

The data manager says similar mergers among exchanges and index providers that have resulted in dominant positions have seen vendors and services become "more expensive and aggressive."

**Lance Uggla**
IHS Markit

> "I don't think this is about them really taking on Bloomberg and Refinitiv, as … in many cases data vendors are specialists in particular areas and often dominate those. This is more about what they can do with a huge inventory of data and distribution, and making it easy for clients to consume."
> **Mark Hepsworth, Alveo**

Douglas B. Taylor, principal of DouglasBTaylor International Consulting—not to be confused with the Burton-Taylor business that is now owned by TP Icap—says the deal strengthens S&P's hold on the investment banking space, and is an intelligent move for customers and shareholders. "Their combined tools will allow customers to make better decisions. But they are unlikely to get cheaper, because if you have all this in one place, you have pricing power," Taylor says.

**Mark Hepsworth**
Alveo

"Markit brings S&P size and pricing power, because it owns a significant piece of the market—and not just in the index space. There is not much overlap, so it's pretty complementary, and creates more value for customers with a broader offering. But, it also gives them a dominant position, so they can maintain or raise prices," says Tobias Sproehnle, CEO of index startup Moorgate Benchmarks, who spent almost eight years at Markit between 2006 and 2014 as head of credit indexes and head of cash bond indexes, and whose business partners Gareth Parker and Mark Pralle held senior index roles at S&P and Markit, respectively.

"Where index providers should have put money into creating new products, they've instead put effort into refining their commercial models to get more money out of existing products," Sproehnle says, adding that index consumers are demanding greater customization as strategies such as ESG gain popularity, but also become more subjective and user-

driven. "Customers are demanding more flexibility, and that will require innovation."

Another long-term bank market data management professional expects the combined vendor's commercial terms to become more rigidly dependent on enterprise licensing models, rather than offering more flexible, metered consumption models.

And a data executive at a European exchange says the added breadth of content creates a one-stop shop for some data types, which could be beneficial to clients, but acknowledges that the merged vendor's control of identifiers and classifications and pricing power over certain datasets could prove "problematic." Since both providers fall under multiple forms of regulation, gaining regulatory approval may also be problematic, this source adds, with the sheer size of the deal potentially attracting regulatory scrutiny.

"I anticipate it will receive more scrutiny than in a usual year because of the intense focus of regulators on concentration risk and competition within the market data space," says Virginie O'Shea, founder and CEO of Firebrand Research, who adds that the vendors may be forced to divest some of these "problematic" assets.

### Opportunities

Despite Sproehnle's concerns about the combined vendor's ability to innovate in the index space, S&P officials believe its analytical tools, including the technology acquired via its purchase of Kensho—which tags unstructured data to make it easier to cross-reference companies and people for use by trading strategies—will help to create valuable alternative datasets from the Data Lake being built by IHS Markit from the wealth of legacy IHS reports and information services.

"There are a lot of complementary assets, and ways we can get data into our Market Intelligence platform, and ways we can leverage [Markit's] Data Lake," said S&P's Peterson.

IHS Markit has already had some success creating new datasets from
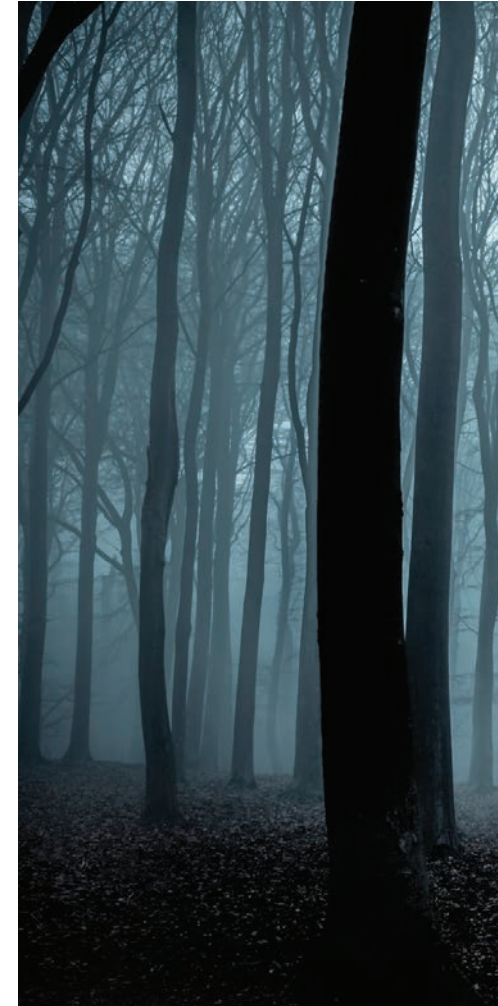
content within its Data Lake, such as a dataset of auto sales and production data, released in December 2018. "Post the IHS Markit merger, we started a journey to build out a Data Lake with infrastructure provided by Amazon Web Services. That journey has been three years, and the Data Lake is in top shape for this integration. It's one of the things our teams are very excited about," said IHS Markit CEO Lance Uggla on the announcement call. Uggla will serve as an advisor to the company for one year after the deal closes.

"When you combine content in an organized way—structured and unstructured content—and add the data science capabilities of Kensho and our teams, you end up with combinations and decision-making tools that are really unparalleled, and it's exactly what our customers want

in this information period, where content is proving so important," Uggla added.

Taylor highlights the combination of Kensho with IHS Markit's data. "With Kensho, S&P has that intelligent data-mining capability, and now it has more data and a tool to help utilize that," he says. "Beyond the obvious asset that the Kensho acquisition brought, I think the real asset was its talent. That acquisition brought some of the smartest minds in the field onto the S&P staff … so now S&P can use them to do all sorts of other things, and now they'll have all that new data from IHS Markit to play with."

## Final Piece of the Puzzle?

Despite emphasizing how the deal rounds out S&P's breadth of coverage, the vendor also hinted at more acquisitions in the future. Announcing the deal, Peterson said the all-share deal allowed S&P to maintain a strong balance sheet and credit profile to allow for "future capital deployment," while CFO Ewout Steenbergen said the new company will generate annual free cash flow in excess of $5 billion by 2023 to use to accelerate organic growth or to pursue strategic M&A.

One area the vendor might consider for its next acquisition could be an execution venue specializing in the data areas it serves, sources say, which could add transaction revenues, create lock-in to data products, and generate new data and index products.

"Looking at the combined business and who they're going up against—LSE/Refinitiv, Bloomberg, Intercontinental Exchange—one of the things that's different is that they don't have any sort of trading platform. So maybe the next step would be to close that gap … if they

**Douglas B. Taylor**
DouglasBTaylor International Consulting

wanted a more complete value chain," Sproehnle says.

And while currently just speculation, it's speculation that fits a trend observed by others.

"There is a massive land grab to become the central platform. It's all about scale and being the platform for full front-to-back trading, from idea generation, to portfolio construction implementation, to trading and execution, to middle-office processing, to back-office regulatory reporting—all of these services combined," says Brad Bailey, research director, capital markets at Celent. "If buy-side firms are comfortable with a single vendor and getting everything they want for security and knowing that these platforms will be open enough if they want to integrate with other applications, this front-to-back perspective will be huge for serving the buy side over the next decade." wt

Max Bowie

# S&P–IHS Markit: It's Not Over Yet

S&P may have agreed its takeover of IHS Markit, but there may yet be some surprises before the deal actually closes. Max ponders what might lie in store for the companies over the next six months.

The Covid-19 pandemic has dulled M&A activity in 2020. However, in September, as the virus raged around the globe, IHS Markit CEO Lance Uggla got a call from S&P CEO Doug Peterson, proposing to acquire the business. By the end of November, a deal was agreed, and unanimously approved by the management and boards of both companies.

I'm no investment banker, but given the scale of both companies, the deal seems to have been settled quickly—especially considering the uncertainty created by the ongoing Covid pandemic. Others agree, noting that both companies appear "highly motivated."

One reason for a quick negotiation may simply be that the deal was such an obvious win–win that little negotiation was needed. Perhaps the timing was right for Uggla, who has expressed interest in other unspecified projects, which may be timely in nature to specific market circumstances. But why—if, as sources say, Markit has been on S&P's radar as a potential acquisition for several years already—act now when Markit's valuation is high? Possibly both took note of the trend of exchanges acquiring data vendors and index providers (for example, Intercontinental Exchange buying Interactive Data, the London Stock Exchange buying FTSE Russell and Refinitiv) and wanted to act before an exchange snapped one of them up, or to create a business big enough to rival an exchange-vendor combination.

On that basis, the combined S&P–Markit still lacks a trading platform or venue component, which some say could be the next step in its growth plan. Certainly, if I owned a trading platform, for example, specializing in fixed income and maybe credit derivatives, I'd be polishing that up in the hope of receiving one of Doug Peterson's compelling phone calls.

Bob Iati, director at TP Icap-owned market research firm Burton-Taylor International Consulting, doesn't see the merged vendor trying to "swim in those waters. S&P and IHS Markit have defined their respective firms as providers of information. Sometimes that does not align well with being a trading system provider." But while Iati says the combined vendor can compete successfully by "staying in its lanes," there is a distinct trend underway among vendors and trading venues, and as more of the remaining players choose that route, options will become limited for those left. The question is, do S&P and Markit want to compete head-on, or carve out a space all of their own that—as Burton-Taylor says in its report analyzing the deal—will "challenge" the current hierarchy of data vendors.

Certainly, the scale of the deal means there are few other players who are in a position to exploit that data, or who are even in a position to absorb a company of IHS Markit's size. Given the missing transactional element, possibly an exchange such as Intercontinental Exchange (which would see it combined with the former Interactive Data business that comprises ICE Data Services) or CME—or possibly a rival index provider concerned about S&P gaining ground through a Markit land-grab, such as MSCI, whose portfolio and risk analytics Markit just integrated into its thinkFolio investment management platform.

And though the companies agree, shareholder approval may be a wild card. The all-share deal is good for S&P because it leaves a healthy balance sheet for more investment and potential acquisitions. But is it good for shareholders, who might entertain a cash bid at a smaller valuation?

The other wild card is how hard regulators and competition authorities will scrutinize the deal, and whether they'll rubber-stamp it, or throw any roadblocks in the way.

Though the vendors have limited overlap, the deal's complementary nature does create areas that regulators might view with concern, such as its reference data and symbology (RED and Cusips)—though the companies may be willing to live with spinning off certain areas if needed to seal the deal. If the speed of the agreement reflects an urgent requirement for data to address a specific market need, any significant delays might make the deal less attractive—as might the prospect of being forced to divest any of the assets that make it attractive in the first place.

And while IHS and Markit both grew by acquisition—more than 120 in total between them—they have limited experience with divestments. Certainly, should that happen, there will be plenty of interested parties waiting to snap up any leftovers—possibly even Uggla, himself, once his one-year advisor term expires. wt

Jo Wright

# EU's Cloud Vision Still Pie in the Sky

The supporters of a plan for a federated cloud architecture in Europe held a conference to discuss development plans, but, says Jo, it's still unclear how the concept will work in practice.

Supporters of an ambitious new project for a Europe-wide federated data infrastructure named Gaia-X held a conference earlier this month to lay out the plans for its development. But actual, concrete details on how Gaia-X is going to work in practice were scarce.

Gaia-X was initiated by the German and French governments. It arose from the sudden urgency around an old concept with a slippery meaning—data sovereignty; essentially, it is an attempt to wrest some share of innovation in artificial intelligence from China and the US, and to dilute reliance in the EU on US- and China-based hyperscalers such as Alibaba, Amazon Web Services, and Google.

Supporters say Gaia-X is conceived as a decentralized cloud architecture based on openness and interoperability. But even after having sat through a two-day conference of speeches and presentations, I can't say that there seems to be more to it than high-level ambitions, at least at this point.

Virginie O'Shea, CEO of Firebrand Research, who also attended the conference, was also left with more questions than answers. "We need to understand the interoperability elements—how are they going to achieve that? Who is setting the standards for this work? What kind of governance structure will be in place across the federated entities? What will the initial use-case be? How will they phase it? Who will oversee the governing body? Where will it be located? France? Germany?" she says.

Even on the level of sector-specific implementations of Gaia-X, there wasn't a lot of detail to come out of the conference. Olivier Sichel, deputy CEO of the Caisse Des Depots et Consignations, the investment arm of the French government, presented on the financial services aspect of Gaia-X. He said "finance Gaia-X" would be underpinned by a platform called the Financial Big Data Cluster (FBDC).

The FBDC's website says it will "integrate the previously unconnected financial data of companies, authorities,

> **I suspect that the EU is setting itself up for an expensive white elephant**

and science in a common data pool, and will be optimized for the development of artificial intelligence applications and systems." Data will be stored in a "secure data vault" that will allow access to data by users in gradations according to their sensitivity, as well as analytics and computing capacity.

Sichel said Gaia-X will offer new opportunities for a more intensive use of AI and data-sharing, which could help with fraud management, terrorism reporting, and anti-money laundering capabilities. It could accelerate the development of new products that improve monetary policy, the oversight of stock exchanges, and enhance market integrity by detecting manipulation with new AI methods.

The use cases that Sichel mentioned are certainly seductive—who wouldn't like to access new AI tools for anti-money laundering processes, for instance, which don't generate revenue and just represent serious risk? But, again, how this will be achieved in Gaia-X wasn't clear from Sichel's presentation.

O'Shea says the finance Gaia-X seems to be "some sort of pan-European shared services cloud-based architecture that will be endorsed by national regulators and governments for use within the sector. The federated concept seems to imply it will be national entities operating interoperable cloud environments rather than one big cloud provider. But the details remain pretty vague."

Supplying a data infrastructure across borders and to all industries, as well as to private individuals; allowing secure access to data, with authentication and access controls in place—well, that seems to me like a rather large undertaking on which to embark without having the rails in place.

It's possible, of course, that there are more concrete plans in development than were made public at the conference. But I suspect that the EU is setting itself up for an expensive white elephant. The European Commission and national governments involved have pledged to support Gaia-X, which "means investment on a huge scale to get this off the ground," as O'Shea puts it, and explains the eagerness of the major cloud providers to be involved with the project.

Without a concrete plan, Gaia-X might benefit no one but—ironically enough—the US- and China-based hyperscalers. wt

# The Front-Office Data Challenge

State Street's Marko Milek says the trend of outsourcing data management is becoming more mainstream.

Agile, dynamic, and real-time are key front-office data requirements. But they can be difficult to reconcile with the industrialized data management and quality certification processes governing the middle and back offices.

Ultimately, the challenge the industry is trying to address is that of putting a portfolio manager in a position, before their market opens, to have all the required information lined up—accurate, timely, and complete—so that they can run the workflows and make investment decisions on behalf of their clients.

A front-office user needs good positions data, but that is not all that matters. Equally important are the nature and context of that position. For example, if a portfolio manager is looking at cash, is it "trade date" cash, "settlement" cash, "good" cash, or is it "still being reconciled" cash? Similarly for reference data—be it traditional market information such as prices, indices, and benchmarks, or more novel sources like environmental, social and governance (ESG), investor behaviors, or sentiment indicators—portfolio managers need to know the source, whether it has been applied correctly, and if it is complete. It's not good enough if only 80% of the portfolio has been enriched with the right reference data sets and 20% of the positions have serious gaps, meaning some analytics aren't quite complete and probably cannot be calculated.

The cash positions, the reference data and the analytics all need to line up. You wouldn't typically understand that so well if you were just looking at it through the lens of a custodian, or even a middle office. You have to be in the portfolio manager's chair, looking at the full suite of analytics at a security and portfolio level, and understanding how those positions are made.

**Marko Milek**
State Street

Having a close understanding of the investment decision-making process also helps when managing time constraints. One of the things we hear all the time is, "I'd like my data to be more timely. I want real-time everything—real-time positions, real-time cash." Adequately managing these challenges requires an in-depth understanding of the investment process itself.

The key questions are: How "real time" do you really need this data to be, and what context do you want around it? Is the current workflow the only way to drive the outcome, or can that workflow be configured differently? Answering these questions requires the perspective of understanding what the portfolio manager is trying to do, or what the algorithm

really matters when there are multiple starts of days, multiple intra-days and end-of-days. Being able to control that data flow at the portfolio level is key as we move away from the once-a-day batch delivery model from five or 10 years ago.

These days, the time distance between data arriving, being normalized and being applied has been reduced from 12 or 24 hours to minutes or seconds.

I'm excited about State Street partnering with major technology providers that have developed great capabilities, such as Microsoft's Azure cloud infrastructure, to continue solving some of these hard problems.

The trend toward outsourcing of data management is shifting more into

> "
>
> The trend toward outsourcing of data management is shifting more into the mainstream. Data-as-a-service is attracting both asset managers and asset owners alike, where the need for timely, accurate and complete data is becoming increasingly important.

is trying to do with that data. Often, upon closer investigation, what managers really need is to have an update before 10 am, or every couple of hours, rather than in real time.

The Asia-Pacific region faces a somewhat unique set of time zone-related challenges that need to be understood and managed. If the core business processes of a data vendor or service provider are geared around North America, then that is going to leave a lot of people in Asia-Pacific markets a little underserved. Also, if you have global portfolios with data needs across different regions, you have to get into the details of what

the mainstream. Data-as-a-service is attracting both asset managers and asset owners alike, where the need for timely, accurate and complete data is becoming increasingly important. They are looking for a data custodian that can hold onto the data, steward it, organize it, deliver it in a number of ways, provide connectivity through APIs, and be the golden source copy.

*Marko Milek is a managing director and Asia-Pacific head of data and analytics at State Street in Singapore. He is responsible for managing and developing State Street's data, analytics, research and advisory functions in the region.*

# Human Capital

### BNY Mellon Taps Fangfang Chen for Senior Role

BNY Mellon has appointed Fangfang Chen as chair and head of asset servicing and digital for Asia-Pacific, with effect from January 4.

Chen will oversee the execution of the company's strategic growth agenda in the region. She will also provide oversight of its regional structure, including regulatory and legal compliance. She will be based in Hong Kong and report to Hani Kablawi, chairman of international, and to James Slater, head of asset servicing client coverage and front-office solutions.

Chen has held numerous leadership roles in the US, China and Hong Kong, including 12 years at State Street. She joins from Algorand Foundation, a financial technology startup involved in the development of an enterprise-scale public blockchain technology.

### Liquidnet Hires Steven Nichols as Head of NLP

Liquidnet has appointed Steven Nichols head of natural language processing (NLP) and unstructured data.



**Fangfang Chen**



**Jenny Chen**



**Kevin Morgan**



**Steven Nichols**

In his new role, he will guide the strategic direction of the NLP team while serving as one of the leaders on the Liquidnet data science team.

Nichols joined Liquidnet through its 2019 acquisition of NLP technology company Prattle, where he was a director of data science. He has helped develop Liquidnet's NLP capabilities and its integration into the Liquidnet Investment Analytics product suite, combining artificial intelligence tools with traditional and alternative data.

He will report to Liquidnet's chief data scientist, Tom Doris.

Liquidnet has also announced the hires of three new data scientists—Nicholas Burtch, Anthony Schramm, and Yusong Liu—to its team.

### Ex-HSBC Data Head Morgan Moves to Data Vendor Evaluate

Kevin Morgan, former head of market data at HSBC, has joined Evaluate, a provider of drug trials and clinical trials data overlaid with financial market data, as chief data operations officer, with a remit to combine its existing data services with technologies such as natural language processing, artificial intelligence and robotics to create new solutions for clients.

Morgan was most recently a capital markets consultant at GreenBirch, prior to which he was a self-employed management consultant, and spent almost nine years at HSBC, first as global head of data operations, then as global head of market data and CRM.

In his new role, he reports to Evaluate CEO Deborah Kobewka.

### Jenny Chen joins as Americas Head of Sales at big xyt

Big xyt has announced the appointment of Jenny Chen as managing director, head of sales in the Americas,

responsible for accelerating growth in the US and supporting clients in leveraging big xyt solutions to transform their trading performance and analysis.

Based in New York, Chen joins from Societe Generale, where she spent eight years as MD, head of global execution services, overseeing program trading, electronic trading, cash equities, and futures.

### DTCC Taps Bob Stewart for Senior Role

The Depository Trust & Clearing Corporation has hired Bob Stewart as executive director of Institutional Trade Processing product management.

In the new role, he will partner with the industry to drive an integrated post-trade lifecycle through settlement finality, develop solutions to help clients prepare and manage CSDR requirements, and focus on the continued adoption of the Alert standing settlement instruction database.

Stewart previously worked at Brown Brothers Harriman, most recently as head of custody product.

### Grigoriy Kozin to Head Business Development at BCS Global

BCS Global Markets has announced the appointment of Grigoriy Kozin as its head of business development for its prime services division.

In his position as head of business development for BCS Prime Services, he will be responsible for spearheading the development of the prime brokerage and securities financing business lines. Kozin will also strengthen the firm's multiple revenue streams and drive innovation within its existing international product suite.

Kozin joins BCS from Sova Capital, where he spent over a decade in the

business development and marketing department.

He will be based in London, and reports to UK CEO Tim Bevan.

## Steve Miele Joins SIP Operating Committees

The Operating Committees of the Securities Information Processors has appointed Steve Miele as a member of its advisory committee.

The 11-member committee represents the diverse viewpoints of the marketplace, with representatives from retail investment, vendors, institutional brokerages, and the alternative trading system (ATS) community.

Miele is chief strategy officer at LeveL ATS, an independent US equity ATS that provides a continuous crossing platform in a highly stable dark pool environment. Prior to joining LeveL ATS, Miele was an equity trader for Fidelity Capital Markets on the Boston Stock Exchange.

## Quandl Hires Hamza Khan as Head of European Data

Nasdaq's Quandl, a provider of alternative and financial data, has hired Hamza Khan as head of European data.

Khan, who is based in Amsterdam, will lead Quandl's data strategy in Europe and help expand its presence in the region. He was formerly CEO and founder of Suburbia, a technology firm specializing in alternative data solutions. Prior to that, he was the head of commodities strategy at ING.

## State Street Appoints Rick Lacaille to Lead ESG Program

State Street has hired Richard Lacaille as senior investment advisor.

Lacaille will provide enterprise-wide leadership of the company's environmental, social and governance solutions, services, and thought leadership across all of State Street's businesses. He will report to Ronald O'Hanley, chairman and CEO of State Street Corporation.

## MACKEYRMS APPOINTS SHERYL HAWK TO SENIOR ROLE

MackeyRMS, a provider of software-as-a-service (SaaS)-based research management software, has announced the appointment of Sheryl Hawk as vice president, customer success.

In the new role, Hawk will lead the firm's customer success team and oversee the expansion of data-driven initiatives.

She joins from Customer Imperative, a customer success consulting firm, where she was a principal consultant and board advisor focused on leading the development of scalable customer success strategy and programs at SaaS companies.

Prior to Customer Imperative, Hawk


**Sheryl Hawk**

led customer success teams at Agari and Whitehat Security—two SaaS companies focused on corporate and enterprise security—and Conversica, an artificial intelligence-driven SaaS company focused in improving lead engagement for sales and marketing organizations.


**Bob Stewart**

The firm has also promoted Lori Heinel to global chief investment officer for State Street Global Advisors.

Heinel joined State Street in 2014 as chief portfolio strategist and has served as deputy global chief investment officer since 2016.

She will oversee the full spectrum of investment capabilities from index funds and exchange-traded funds to active, multi-asset class solutions and alternative investments. She will lead a team of more than 600 globally.

Heinel reports to Cyrus Taraporevala, president and CEO of State Street Global Advisors.

## Mario Platt Joins CloudMargin as VP, Head of Info Security

Mario Platt has joined CloudMargin as vice president, head of information security.

Platt has previously provided information security and data protection consultancy services to a variety of UK-based organizations through Privacy Beacon, a firm he founded in 2018. He has also work in security-related roles at Dixons Carphone Group, Three UK, and Vodafone UK.

## Mediant Announces Chris Nobles as CTO

Mediant, a provider of investor communications technology and

technology-enabled solutions to banks, brokers, corporations, and funds, has appointed Chris Nobles as CTO.

In the new role, Nobles will oversee the ongoing expansion of Mediant's technology platform.

He joined Mediant at its inception in 2002 and led the development of broker advisor and investor user interfaces at the firm.

## Greenwich Taps David Easthope for Senior Advisory Role

Greenwich Associates has appointed David Easthope as senior advisor for market structure and technology.

Easthope will lead the team's efforts in capital markets fintech, market data, and analytics. He joins Greenwich Associates from Celent, a research and advisory company, where he led the firm's capital markets practice.

## GLMX Appoints Lauren Carroll as General Counsel and CAO

GLMX has appointed Lauren Carroll as general counsel and chief administrative officer, responsible for providing strategic legal advice and direction for the global business.

She joins GLMX from MarketAxess, where she was general counsel for LiquidityEdge, the MarketAxess rates business, and assistant general counsel. **w**t

# SmartStream

# Let's talk about the next wave in AI, Machine Learning & Managed Services

SmartStream's fully integrated suite of solutions and platform services for middle- and back-office operations are more relevant than ever – proven to deliver uninterrupted services to critical processes in the most testing conditions. Their use has allowed our customers to gain greater control, reduce costs, mitigate risk and accurately comply with regulation.

With AI and machine learning growing in maturity, these technologies are now being embedded in all of our solutions and can be consumed faster than ever either as managed services or in the cloud.

Stay tuned for some major announcements or simply book a meeting to find out why over 70 of the world's top 100 banks continue to rely on SmartStream.

**info@smartstream-stp.com**
**smartstream-stp.com**