STATE STREET GLOBAL EXCHANGE

From big data to smart data

In an age of vast amounts of information, volume doesn't mean value. While the sheer quantity of data promises new sources of knowledge, the challenge for the buy side is to use information to make smarter investment decisions. In a webinar forum convened by *Risk* and sponsored by *State Street Global Exchange*, a panel of experts discusses how they are adapting their approaches to turn big data into smart data

Risk: Big data means different things to different people. What are the buy side's priorities in this space, and how are needs for data changing?

Jeff Conway, State Street Global Exchange: To better understand the challenges our clients were facing and what these words meant to them, State Street commissioned a survey from the Economist Intelligence Unit, covering more than 400 institutional investors across client segments.

Ninety per cent of respondents saw data and analytics as a strategic priority, and just under 40% saw it as their highest strategic priority. As we looked into how much they were going to invest in that space over time, 86% said they were going to spend more in the next three years, and one in 10 said they were going to spend in excess of 20% more. Ultimately, what we learned from the survey was that data and analytics capabilities create a competitive advantage – two-thirds of respondents said that they would be a key differentiator in the future.

We translated that into a few areas where we are focused on driving change. One was around new regulation, and the pace of it. The ability to master regulatory complexity is seen as a competitive differentiator, which you probably wouldn't have said pre-2008.

The other area is around risk. Nearly 80% of executives believe risk is the highest priority on management agendas today. That's not surprising given everything that's going on from a regulatory perspective. If you compare that to 2007, 30% of respondents felt the same way. We're seeing changes in the risk management space, where buy sides are adopting sell-side risk management practices and are connecting risk with performance.

Vijay Nadendla, Fortress Investment Group: If we ask an analyst to marry model risk and actual data, he's going to put it into Excel and manually make the connections, which can lead to 'Excel hell'. Excel is the 'small' big data providing velocity, rapidness of integration and varied functionality. You need to build a system, not to give static reports, but to mimic the characteristics of Excel, i.e. you can put in any data you want and you can copy, modify or delete it. You can take the data, make a new spreadsheet, add your model data to it and create a new scenario to produce results. You can bucket and create new variables, name lists and calculations.

THE PANEL (left to right)

Vijay Nadendla, vice-president of enterprise data and applications, Fortress Investment Group
Klay Stack, managing director, Marathon Asset Management
Jeff Conway, executive vice-president, State Street Global Exchange
Edwin Amaya, director, investment strategies & solutions, MetLife

At Fortress, we set out to build a data warehouse that had those characteristics. We were able to abstract out and create scenarios, to make multiple user-defined hierarchies and do all the calculations. We have the ability to put the data in whenever we want, quickly, or freeze it for analysis stability. That, I think, is the ultimate essence of big data.

Edwin Amaya, MetLife: We apply an open architectural model, where we engage in sub-advisory arrangements with asset managers to manage assets for our portfolios, which in turn have different complexities given some of the guarantees involved under the insurance solutions. The challenge of big data today is the level of data quality, as well as quantity, you can collect for the purposes of making smart decisions. And that's about trying to gather as much information from as many sources as we can, including the sub-adviser, who is responsible for trading the assets on your behalf, but also internal sources such as custodians or proprietary databases in-house. We need to collect all this information, understand the underlying risk of the portfolios and decide what is best for the portfolios to manage going forward.

We're always trying to understand several things – capital appreciation versus capital preservation, systemic versus idiosyncratic risk. We want to understand what scenario could possibly occur that would affect the assets in these portfolios. That's a challenging thing to do these days given the complexity of the market, particularly in fixed income, which has become very broad over time. The underlying risk of each security, or for each subsector, has become very separate or diverse. We try to understand not just the data coming in, but also the content of that data and what it tells us in terms of the underlying risk.

Klay Stack, Marathon Asset Management: It is about looking at opportunities that might not be mainstream, and that could be in the fixed-income or credit markets, whether it's sovereign debt or bank debt – things such as structured credit and structured debt. It is looking for an opportunity – an opening – taking advantage of that and moving quickly on it.

Big data is particularly relevant to us in the residential and commercial mortgage-backed securities (MBS) space because we had to build a decision analytics platform for that trading desk and they're very data-driven. They are looking at credit rating information, home sales information, estimated and actual public record information, dealer inventory, portfolio information and bids wanted in competition. They've got all this data they want to bring together and correlate. That's how managers make their decisions – do we have a particular position in a particular tranche or security, and do we want to take a bigger position or do we want to get out of it? How do you make those decisions? Correlating all these data points is a big part of it. So, you have to be nimble, and building that kind of warehouse enables that kind of nimbleness.

Risk: When discussing big data, just how big are we talking about? And what are the challenges of handling data on this scale?

Klay Stack: Big data is a relative term. For someone like MetLife, big data is much bigger than for us. For our firm, we're talking in the five- to 10-terabyte data range and you need to have infrastructure underneath that supports that and scales it. If you've got the infrastructure, you're well prepared for bringing in more and more data points. The marketplace has progressed quickly and one of the main drivers is that computing storage has gotten cheaper, so more people can now do more things. If more data is coming, I can do more with it; I can process more, calculate more, and derive more data from that. We're constantly driving towards that scale to make sure we can take any data points we need.

Jeff Conway: The overwhelming factor here is the volume, the velocity and the complexity of data. It's a highly complex challenge to aggregate, normalise and then translate data positions into risk decisions – and do it at the pace required in today's world. And, in the next three years, that problem is going to expand. When we think about investment strategies and solving problems, we have a habit of solving the problem for today. But, given the pace of change, I wonder if we are prepared to solve that problem three years on, unless we do something fundamentally different.

Jeff Conway

Klay Stack: Yes, I think that's the challenge. It's anticipation; trying to anticipate where the market's going, where the business is going. IT is usually lower on that decision-making chain, we're not directly involved in investment decisions, so we have to trend it ourselves. Sometimes an innocuous question asked by a portfolio manager suddenly turns into an investment strategy. So I'm always listening for those little telltale signs.

Vijay Nadendla: A larger volume of data doesn't mean you store everything. You have to devise techniques to store gigabytes out of terabytes. This includes creation of multidimensional aggregations where you are only bound by the cardinality of the space, not the volume of the data processed, and abstraction where both the volume of data and the cardinality of the multidimensional space is reduced.

Risk: Has there been an increase in use of big data by asset managers, and if so, to what extent?

Edwin Amaya: It has been massive. And it's necessary given today's market. The complexity of the investment marketplace has changed dramatically. It is important we understand the risk/reward of any investment vehicle being put on asset portfolios, not just in the short term but also the long term, because, as an investor, we'd like to think our investment horizon is much longer given the product construction that's involved. The challenge lies in that we want asset managers to be as transparent as possible. So the question is, are you getting good-quality data that allows you to make smart decisions? I think that's an ongoing process. It is an open discussion that you have with your portfolio advisers. Is the information that you're gathering, that you're analysing in-house, giving you want you need to make decisions going forward?

Risk: Are there any gaps in big data? Is there an area where you see room for improvement, for example, in the quality of data users receive?

Vijay Nadendla: There is a gap in the processing of data – not in the data itself. We need to separate data from the metadata. Data elements are going to be the same. But what's going to be different is how certain people interpret that data – government, customers or the business all have different views. We need to separate the physical data of what happens in the markets from the layer you want to put on top of it.

A lot of tools have been developed over the past 20 years for physical data integration. You can get disparate data sources and put them together. But data has content, context and completeness. You need to know the context, and the context is not always with the data. Somebody needs to logically put it together. That is the harder part, and it's a human that does that.

Edwin Amaya: Data collection has to have a thesis, and that thesis needs to be evaluated, tested and defined. Anybody can collect data but, if it doesn't help you reach your objective in terms of what you're trying to define, evaluate, test and ultimately make decisions on, it's no help to you at all. I think that's an ongoing discussion you need to have with your business partners.

Klay Stack: Data is looked at in a lot of different ways from the business side. The way attribution is done from an investor's point of view or a regulatory point of view is much different, in terms of the way they want to see data points. You change the way you're handling and storing attributes because, for example, the way one defines liquidity might be different from the other, the way you categorise asset class and sub-asset class could be different. That's a challenge we have to keep up with, and it just adds to the big data piece of it.

Edwin Amaya

Klay Stack

Risk: Like other market sectors, asset managers are facing a wave of new regulation. How have data and analytics needs changed in response to this?

Jeff Conway: Regulation is driving at higher rigour around risk management. So now you get into scenario testing and stress testing, which creates even more demand on the data management process. But it also leads to new and more factor-based analytic models. So data regulation is fundamentally changing how we're looking and managing portfolios.

Thinking about alternative regulatory reporting such as Form PF or the Alternative Investment Fund Managers Directive or even Solvency II in the insurance space – the complexity of reporting is very high. It's not just about getting your custodian data and reports off the custodian records. The ability to integrate multiple sources of data and make a representation around your portfolio set is highly complex, and it's a requirement that is fast upon us.

On the back of this regulatory change, there will be real data and analytic stress on meeting regulatory demand at the pace required. But there's also a whole level of uncertainty that will occur because the markets will continue to change in reaction to that regulation.

Edwin Amaya: Indeed, there are regulatory changes that will occur and we all need to understand the impact. In some cases, the US Dodd-Frank Act is an ongoing process to be determined. There are other situations coming to market as well that we need to understand, such as the Treasury Market Practice Group's guidance on MBS trading.

In dealing with asset managers, communication is important to understand not just the operational impact, but also the fundamental impact of regulatory change on philosophy and implementation. What does it mean for asset managers when they have to be concerned about how much collateral they need to post? We need to understand what portfolios will look like structurally as they continue to be managed going forward. And that only happens through data collection and communication.

Klay Stack: It has certainly added to the big data challenge. You must have a certain amount of competence around the regulatory reporting so you don't even have to think about it, because you don't want to be distracted by formulating reporting every quarter or every month. You need rigour around the process of data integrity, whether it's procedural referential data integrity checks at the table level, for example, or establishing the right foreign and primary keys. And probably most important is the controls. You need to have a discipline to enforce the controls, because you could make these correlations to ensure your data's valid but, if you're not following your controls and you don't have a rigorous procedure in place, then you've got bad data.

Risk: What you are doing in terms of investment in data mining, and optimising your advantage from big data?

Klay Stack: The investments are in infrastructure; you need to make sure it is always a step ahead of where the business is going to be, because they're constantly bringing in data. On the product side, we're investing our time into looking for tools, whether they be third-party tools or tools we can create, to help us refine data and perform the analytics. There's always the need for that human element, whether it is a data steward or master of the security data, to make sure the story of the data is being told logically. You need the human element to be able to perform the analytics to assess correlations and make investment decisions. For us, the investment is making sure we have the right analytic tools to give them the power to do that.

Risk: For some, the challenge is to assess how managers are making investments. How do you set a standard for that?

Edwin Amaya: There has to be a process in place first of all, to understand what the objectives of each portfolio are. Ultimately you need the right technology to allow you to get the data you need to reach that objective. Not every vendor will be focused on your particular strategy, so you have to try to find the right vendor that can do that for you. Data collecting has to be associated with the objective you're trying to reach, in terms of the underlying risk. We have portfolios that are highly diversified and some that are very sector-specific. What you're trying to use might not match every resource available to you, so it is often a question of what you have in place collectively that allows you to run independent analysis of what the risk will be. That's a big challenge when dealing with asset managers that have different implementations and different philosophies.

Vijay Nadendla: You need to have people with 'data sense' to analyse the data, find where the breaks are, and be able to understand and put it all together. You also need time and space – you can't just increase the number of staff. It's going to take time to understand and get this all done.

Our system wasn't very expensive and it wasn't a huge group, but it was a small group of highly competent people who had the space to build this, and they were able to achieve it very successfully. With the new system, the number of reports that analysts made increased exponentially. We were asked by different divisions in the company to implement it in their areas, so a tool built for one space – let's say accounting – got implemented in areas such as asset management, risk and fund accounting. Implementing that not only took money, but also people, time and space.

Vijay Nadendla

Risk: The growth in use of big data looks set to continue, but firms could struggle to make the investment to keep pace with that. What changes do you expect to see in big data in the years ahead?

Klay Stack: One of the big drivers will be that technology that was once reserved for very large firms will become more accessible for smaller firms. Years ago, at Marathon it would have been a challenge to do high-performance computing and grid computing, but now we can. That's one way we've been able to scale our process and applications. Some big-data science out there is still reserved for the larger firms but, as technology gets more refined and smarter solutions come out, and also just by sheer volume bringing the price point down, this technology becomes more accessible. Technology is a bit cyclical – from mainframe to desktop, back to centralised computing with virtualisation – so there might not be any new technologies, but there are going to be innovative ways of looking at existing technologies.

Edwin Amaya: Data collection is going to exceed that of prior years, no doubt about it. Having easier access to data will be key for providers going forward. The large effort being made to move things to the cloud is going to make it a lot easier for folks like us to easily gather information. But, of course, the important point for vendors is data quality. There will always be an effort being made on everybody's part, not just significant players but small players as well, to try to get the best pricing possible, but to also get the best vendor possible to meet your needs.

The challenge is creating solutions at a time where customers are looking for customisation. Not every client is the same. As providers adapt to change, they must understand the needs of individual clients; a cookie-cutter model may not necessarily work in the future.

Vijay Nadendla: My vision is that the big guys will find new ways of staying ahead in the technology race, and the small guys will find innovative ways to make do with what they have. I expect data processing to change. Right now, we are limited in certain ways by the hardware and the operating systems. I see the development of programming languages to enable data-set processing rather than one record at a time. At the moment, even our data-set algorithms underneath process only one record at a time. If somebody could come up with systems that process sets of data deep down, that would be a major win. To be able to process huge amounts of data would be a game-changer. There is a lot of inefficiency in trying to process everything one at a time instead of a million at a time.

Jeff Conway: One thing that came out of our survey was that the ability to retain and attract talent is fundamental as we start thinking about this changing landscape. Those that don't make fundamental shifts and invest for the future – and that continue to do things on an incremental basis – are going to find it hard to retain talent. Smart people are not going to want to continue with manual processing or cumbersome data management. We found that the leaders in data are going to be in a much better position to attract and retain talent.

At State Street we too have to adapt, and not incrementally, but by creating a multi-year view of where we need to be. We look at our production and our own technology infrastructure, and we're embarking on a multi-year digitisation effort. We have a significant investment under way to digitise a lot around operational activity. That means the nature of roles we hire for changes, because it's less about the production and more about the data on the back of digitisation. We're also providing a different level of capability to our clients through the launch of State Street Global Exchange, which will allow us to make better investments for the future with a much more informed view of the landscape.

Any time you're dealing with the amount of change we're seeing, asset managers have to make choices. They need to pick their partners wisely and be clear about what they're asking from them. That's going to be fundamental. Many asset managers, whether large or small, are not going to be able to do this on their own. They'll need to choose partners, technology or service partners, to help them through this.

>> View the full proceedings of the *Risk*/State Street industry leaders' debate, *From big data to smart data*, at *www.risk.net/2320946*