How to build an Effective Data Management Strategy in the era of Big Data Technology

Leveraging the latest data management platforms to meet the pressing needs of financial services institutions, in the current regulatory environment.

A practitioner’s view by Simon Trewin, Senior Managing Consultant, Kinaesis.
An effective enterprise data management strategy is made up of a series of pillars. Each one is key to the other and the combination of them together, builds a strong data driven and analytically superior organisation. It is our experience that in some organisations these pillars are not being driven cohesively and the benefits are therefore being wasted. The regulator through BCBS 239 has more or less laid out the ideal. This principle based document highlights the pillars for the solution, however, it does not mandate the path. With additional regulations such as MiFID, CCAR, Volcker, etc. the danger is that organisations kick off programmes without a fully thought through, cohesive path to successfully delivering the holistic data management strategy. To achieve success governance, data quality and data aggregation need to be integrated with the overall project process, and solutions architecture. Without this joined up strategy, the investments made will be squandered, and create additional ‘Run The Bank’ (RTB) cost base. With limited budgets and difficult trading environments, it is hard to justify a great deal of expenditure, on what many consider to be somewhat ethereal concepts, such as data governance. It is therefore important to quickly ascend the maturity ladder, learn fast and improve, and deliver tangible returns from the regulatory driven targeted investment. This paper lays out the roadmap for success, built on the pillars of a good information strategy and it highlights the steps to make improvements, and the pitfalls to avoid.

The Challenge
Looking across this industry there is a lot of legacy baggage, making many established infrastructures difficult and costly to change. A lot of the headwinds are embedded in existing process and culture. This magnifies the challenges of being able to make real change. However, financial services firms cannot be complacent - how long is it before a really disruptive event occurs, where a new bank without this baggage, is able to use the latest technology to get to market at significantly lower operating costs? It has happened in the retail sectors, in travel services, and it is ripe to happen at some point within financial services.

A summary of some of the challenges facing financial services relating to data strategy
How do you eat the elephant of significant and lasting change, whilst delivering the regulatory driven requirements and driving towards greater efficiency? How do you optimise a complex and large global business? In the past, it was less necessary to aggregate information - balance sheets were large and leveraged, regulators accepted siloed information and rarely cross checked. Institutions now face a race against time to pull their enterprise data together to be able to remain competitive and to answer the current regulatory challenges.

The Opportunity
The Kinaesis view, is that you build an information management strategy that embeds into the mind-set of the organisation. You use facts and figures to analyse the effectiveness of your organisation: “What is our return on equity for a trading business”, “What is the profitability of our customer base”, “Is there a better way to be using our balance sheet to server our customers”, “What is the next best action for this customer”. Importantly, you ask these questions in real time, drilling from one question to
another, through responsive and comprehensive analytical tools. This insight is available within the data already available in the organisation. Once it is brought together, an organisation can cut its cost base sensibly, improve revenues, improve customer experience, and achieve competitive advantage.

**The Strategy and Model**

To address this challenge, the data strategy needs to be looked at holistically. Firms need to first assess where they are with regards to each pillar and they also need to assess where they want to be. The combination of outputs of this analysis, provides a foundation for an effective data management strategy, where effort becomes cumulative and every next step brings improvements as part of the overall culture. Requirements need to be driven from business outcomes and then translated into targeted innovation and organisational change. A partnership between leaders within the business, projects and technology needs to be made, as only this combination will enable projects to succeed. One of the common problems that we see is that the businesses want this without any change, they either expect IT to walk around with magic wands and silver bullets to bring back the good times, or they hang onto their existing processes, and expect IT to dance around trying to keep everyone happy. The result is failed projects, or projects that cost huge amounts, or that overrun. It is therefore essential that a good target operating model is established to support the culture.

The next step is to establish an information management change programme, based on where you want to be for each of the information management strategy pillars. Kinaesis focuses on the following 6 pillars: Architecture, Data Delivery, Data Lineage, Data Frequency, Governance and Project Delivery. Getting these to the level that the organisation desires, will provide the foundation for addressing all of the diverse needs of the organisation in the face of the regulators and the demands of the market.

**Architecture**

**The Challenge**

The challenges that exist for a new architecture are greater than ever before. Regulations are demanding the analysis of more data, joined across typically separate data sets, with the requirement that all comparisons and correlations need to produce the same results.

Time is playing a big part as data continues to move, i.e. what is the view of our data at this point in the trading day?

Or when we ran the risk processes? Or the trader sign off? Or the month end accounts? The architecture needs to have temporal consistency to be able to answer these questions.

The demand for a single version of the truth, with many perspectives is required to support consistency throughout the organisation. The ideal scenario is that there is a single copy of data which supports the traders, risk, finance, operations, compliance, MIS, treasury, credit risk, and liquidity. This then demands that data is mastered once and used by multiple stakeholders.

Finally good practice says that this architecture is quick to change, and be dynamic. Reading the BCBS 239 directive, it is mandated that new risk parameters need to feed through into the system rapidly and be integrated into the analysis.

**The Pitfalls**

Traditional data warehouses built on older methodologies and techniques are challenged when it comes to satisfying these demands. Firstly scalability, the level of data and the temporal nature causes considerable versioning problems. The traditional slowly changing dimension logic, starts to constrain your perspectives and temporality. The reprocessing time of the data can eliminate your dynamism. Supporting the many perspectives forces the typical ceilings in analytical schemas to be exposed, or blown away. Taking the wrong approach with the wrong technology leads to a world of pain. Endless performance tuning, slow change cycles, frustrated users, and stressed IT departments.

Traditional architectures through their implementation throw data away, or only store it when it has been modelled, or rely on modelling the entire enterprise before anything is even built. Only storing data that is modelled is initially efficient and allows something to be built quickly, however, once the initial solution is established, change cycles can become prohibitive. Modelling the whole enterprise before moving forwards, leads to huge lead times in projects before benefits are felt and in a climate of limited budgets, there is a very high risk that patience runs out and support fades before it is complete.

**The Solution**

The correct strategy for sourcing is to collect as much data as it is possible from source in one go and storing it without modelling. This gives the benefit of reducing the amount of round trip back to source, saving time and money, and builds the foundation for user driven analytics. Modern low cost storage and processing architectures provide the

Simon Trewin  
Senior Managing Consultant  
Simon.Trewin@kinaesis.com  
Kinaesis.com
means to store and process this data efficiently, when required.

The semantic model and supporting meta data provides the key to this architecture not becoming a free for all and a data wild west. The solution needs a new age of design patterns around data models, using some clever big data techniques, but also needs to pay due respect to the guiding principles of good data management. The models by their nature, become less bound by the direct relationships to support the complex use cases. Instead, integrity is maintained by strong, integrated governance. The governance of the data enforces the principles that enable the insight to be gained. The semantic model will provide a consistency of analysis and the quality of the results, which enables decisions to be driven from the information captured.

The diagram above shows a modern data analytics platform

By using these new design patterns, the data can be stored as a single version of the truth. Now this needs to be presented up to the user to report and analyse the data, through multiple perspectives. To achieve this, a pragmatic solution has to be sought. Multiple marts need to be built and maintained consistently on a powerful analytics platform. A design pattern for this not only gives you the ability to support many users, but it also implements analytical marts as a disposable commodity, to enable ad-hoc analytics and user driven reporting. The technology needs to provide linear scalability and good parallel performance for multi-user concurrent load and query. You also need a simple way of creating new analysis which also needs to be equally scalable. Some of the new SQL tools integrated with Big Data, enable this to be achieved and support a SQL interface that is key to the integration to presentation tools, and usage by a large pool of development and support resources. Finally, you need an infrastructure that gives you a cost effective way of keeping your data.

The combination of these architectural designs will enable the challenges of the current requirements to be solved. Implemented correctly, by using the latest test driven approaches, the solution satisfies the ethos of the BCBS 239 directive and it provides a foundation for the data driven financial services organisation of the future.

Data Governance, Lineage and Quality
As explained above, to achieve the journey to a modern data architecture will require a number of the other data management pillars to be addressed. Governance, lineage and quality need to be strong and embedded in the architecture and the project process.

The Pitfalls
With an architecture that does not explicitly enforce relationships (for the reasons described above) it becomes critical that standard good data management principles are enforced. Without these, there is limited actionable insight to be gained. The governance needs to be integrated into the architecture and the project process. Typically, we see financial services organisations implementing standalone governance solutions, often through advice from external "data experts" or consultancies. These exercises generate a huge set of documentation, which takes massive amounts of time and money and which is out of date, a second after it has been produced.

The Solution
The governance solution needs to lead to the mastering of reference data and the consistent definition, propagation and use of hierarchies, calculations, business terms and business rules. It should be embedded into both the architectural solution and into the defined project process. This will lead to an operating model where the governance solution becomes inherent in the culture and the spirit of the organisation. This in turn, leads to consistent understanding of the purpose and the benefit of following robust and disciplined practice. It is our experience that when this is achieved, then the veracity of insight is accepted as a given and taken forward to being actionable, which leads to increased opportunity for growth and a reduction in cost base.

To embed data governance into the architecture, you need an integrated meta-data driven mapping of your business
the temptation is to jump to solutions before the problem is
pressure to
market is the need to jump to solution too
A second common pitfall that we s
maintenance and leading to poor data quality.
that
of the architecture causing maintenance headaches
scale,
perceived benefits rapidly turn into solutions that do no
wrong foundations
reports driven by users.
Kinaesis
Spotfire,
immediate short term effect.
and
pretty
The toolsets available on the market now that present
pictures to a captivated audience, are very enticing
and can quite often be implemented successfully to
immediate short term effect. Tools such as Qlik, Tableau,
Spotfire, etc., are prevalent throughout the industry.
Kinaesis’ experience is that these are great tools really
enabling that capability of dashboards in seconds, and
reports driven by users. It is also our experience, that if the
wrong foundations are built to support these tools, then the
perceived benefits rapidly turn into solutions that do not scale,
that embed business logic into non repeatable layers of the architecture causing maintenance headaches and
that implement many versions of the truth, increasing cost,
maintenance and leading to poor data quality.
A second common pitfall that we see throughout the
market is the need to jump to solution too quickly. The
pressure to produce results and show progress is huge and
the temptation is to jump to solutions before the problem is
even properly defined or understood.

The Solution
The key to success on these projects is to use a
methodology that clearly articulates business outcomes and
then tracks these through the project process, to the
solution providing checkpoints and controls at each stage,
to bring risks forward to the start to be managed effectively
and to eliminate unknowns.

Agile
Typically, the desire is to run these projects in an agile way,
short sprints, regular deliveries and a user driven and test
driven approach. To get to this place it is important to
realise that elements of the project will only run according
to the waterfall project process. This needs to be taken into
account and the impact of these waterfall elements need to
be identified and minimized. In our experience, sourcing
and infrastructure form the basis for the waterfall. Lead
times run into months.

The diagram above shows the agile analytics delivery
methodology supported with a underlying framework
Commodity virtual server estates with an embedded
capability to scale up and down can be used, to reduce
infrastructure lead times and dependencies. This makes the
case stronger for big data solutions which run on virtual machines.

Implementing a ‘consume all data’, or alternatively
‘consume available data’ approach to sourcing, removes
many of the sourcing dependencies. It also points to the
need to have plenty of landing space for the incoming
source data and the need to make sure that this space is
tought through in relation to processing.

Data Driven
The cost of building a data solution from source, through to
report, is significantly more expensive than to prototype
and profile the data to the solution. It is important to realise that, most likely, the initial specification of reporting does not accurately reflect the desired outcomes. It is not common that users have clear idea of the art of the possible when it comes to information management. This makes scope creep or change inevitable. Using wireframes and story boards, or mocked up analytics in a sand pit environment brings this forward in the project lifecycle, allows stakeholders to understand what they will get prior to commitment of all of the budget to the implementation, eliminating potential waste.

The data driven approach enables the data quality issues within the data to be established early in the project. This enables them to be articulated to the stakeholders and helps manage expectations and decide on actions to take through transparency.

The amount of infrastructure and the capabilities of a data system should be driven through a scientific approach. The algorithms and internal data flows and models can be established using a set of rules, or proofs of concepts. The key thing to understand is that the problem needs to be described properly by defining it accurately and early within the project process to plan for success. Once a framework architecture is established then it needs to be tested to the point of breakdown. This establishes the points at which further investments or changes will need to be made.

The whole of the process above needs to be drawn together with an end to end methodology that provides the confidence, through artefacts and traceability, that the critical business objectives will be met. If implemented correctly then the project artefacts will form the documentation required to support the governance of the data and the platform. The project process should become repeatable and efficient, leading to greater productivity. The results of the process should lead to solutions that meet the expectations and needs of the business. A key outcome of this will be the coordinated teamwork between business, operations and technology in owning and managing the overall solution.

Conclusion
Implementing an effective enterprise data management strategy is made up of a cohesive and complementary approach to each of the pillars of good practice. Point solutions and trying to get to the end goal without going through all of the stations does not work. Building an information driven business is going to take new age data technologies and techniques and an architecture of the future that is scalable, with high performance. The Data governance solution needs to be embedded into the architecture and the project process, to enable it to be implemented and supported in the future. The project process needs to bring everything together to drive the best outcomes for the business and the greatest productivity with reduced risk.
How Kinaesis can help customers with their enterprise information management strategy

At Kinaesis, we combine expertise in high scale data management architectures, metadata management and data analytics, with proven information architecture methodologies and deep experience of risk data and finance. From assessment of where our customers are, against the Kinaesis Enterprise Information Maturity (EIM) Model, to delivery based on the Kinaesis Best Practice Methodology, we use out of the box components and templates to accelerate customer’s projects and ensure value delivery. Kinaesis offer proven data services to accelerate delivery and to reduce the risk and cost of your programme. This includes metadata management (data governance, modelling and lineage), rich data management (data lifecycle, state, temporal design) and data insight (data visualisation, metrics and analytics).

- Our highly experienced managing consultants can get your data management projects and programmes moving forwards successfully.
- Our thought leading solutions architects can turn your vision into a clear defined achievable outcome.
- Our experienced data architects model, transform and organise your data in the latest tools and technologies to provide concrete robust and highly performant solutions.
- Our business data analysts can structure and manage your governance, improve your data quality and help support and improve your project process.
- Our consultants can advise and support your internal staff or we can implement robust and best practice data architectures.

To speak with one of our experts contact us on 020 7347 5666 or info@kinaesis.com

Biography

Simon is co-founder of Kinaesis and Senior Managing Consultant and brings over 21 years’ experience in risk, data, business and technology, gained within the financial services industry. His work has involved a number of roles at top institutions in the City including UBS, HSBC, BNP Paribas, Societe Generale, Credit Suisse, Citi, RBC Capital Markets, and ING. Specialising in front office, risk and P&L control technologies calculation and aggregation of complex risk technology problems. He has taken on various roles from Front office business management to running major technology change / development programmes.

About Kinaesis

Kinaesis are the leading independent consultants in delivering high performance data architectures, enterprise information management and on-demand analytics solutions for Financial Services. We help our customers tap into huge volumes of complex data, unlock its value and bring agile decision making and real-time insight, into key business processes. We are specialists in complex, high-volume, global environments, servicing the extreme analytics and reporting needs of many thousands of end users, based on terabytes of data and billions of rows per day.