



How to deliver your data strategy

by **Paul Hewitt**, Director and Practice Lead for AI, Machine Learning and Big Data, Luxoft

The financial services industry has always relied on data and accurate record-keeping. In this article (based on my Data Management Summit keynote), I'm going to look at the latest data trends and how they're affecting (or about to affect) financial organizations like yours.

1. Data volume

A decade ago, former Google CEO, Eric Schmidt, commented that “every two days we produce as much content as was produced by all of mankind for the 20,000 years before 2003.” Today, the rate of (primarily unstructured) data creation is far, far greater.

Accordingly, to manage the exabytes of big data created daily, we need better tools — particularly around automation and the cloud — because there's way too much data to cope with manually. Automation also involves artificial intelligence (AI) and machine learning, which help organizations automate intelligent decisions based on data.

2. A changing society

COVID-19 has been a significant contributor to societal change, a development that has impacted data too. Our response to the pandemic showed that remote working is possible on a scale that would have been inconceivable a decade ago. Now we have the networking, bandwidth, tools and capabilities to make the disparate workforce a force for good.

For many office-based workers, going virtual didn't change the basics of their working day too much. We live our lives digitally anyway, generating more and more data each day. And as we spend more of our lives online, data privacy becomes a priority. Privacy and security have to be built seamlessly into all of our data tools and technologies. Also, the increasing importance and visibility of data are encouraging the regulators to pay more attention.

3. Change acceleration

Along with the increase in data volume, escalating technological advances are driving new and more extensive transformations in organizations.

Here, we're not just dealing with vast and monolithic datasets; there's a hoard of smaller and more detailed sources of information in so-called “small and wide” datasets. Again, this drives the evolution of flexible tools and designs to cope with big monolithic datasets and small, wide ones.

To master accelerating change, we need to automate data management itself, deploying metadata tools that can help us manage data at scale (e.g., data cataloging, data lineage, etc.).

In fact, we need to automate as much as possible. Some elements are still quite hard to operationalize — AI, machine learning and model management in particular. Although there are many advanced tools, it's still not the easiest part of the process to automate fully. So, this situation drives a need for both standardization and flexibility.

Data should drive the tools, not the other way round. For example, in the early 2000s, the data industry dealt with highly structured SQL databases, which brought a certain rigor to the way we collected the data. Our approach was, “the data we ingest or create has to fulfill the following needs and requirements, as per this predefined schema.”

Organizations don't dictate the structure of the data so much anymore — the data itself dictates the form. So now, we receive tons of largely unstructured data, which we need to figure out what to do with and determine how to extract the information and value from it. Consequently, data tools are becoming more flexible to help us achieve this.

4. Ubiquity

Data is everywhere. It seems like almost everything can generate the stuff — your doorbell, your bicycle, even your running shoes.

The processing of all this data can now occur just about anywhere. With smart devices, the IoT, edge networks, containers and APIs, increasingly, it's more practical to process the data wherever the data sits.

Therefore, we shouldn't tie data tools to a particular location. Some years ago, this resulted in the development of various container-related technologies so that users could process anything anywhere, relatively easily. Today, this drives us toward using fabrics — distributed and interoperable collections of tools and services — rather than one specific tool or cluster.

Drivers of data transformation

Organizations make the data management transformation journey for a variety of different reasons.

Increasingly, **regulatory requirements** such as GDPR are driving data system improvements, with substantial penalties for failure to manage data correctly.

Productivity — aiming to put data to work — is another prominent driver. **More than 60%** of enterprise data goes unused for analytics, creating a gap between potential and actual business insight. In many machine learning proofs-of-concept, far more time is spent interfacing with the correct data than doing valuable work. So, anything we can do to improve data productivity, whether by improving data lakes or enhancing AI-based data interaction, will help the bottom line.

Governance is an increasing area of interest. It's vital to make sure we've got a good grip on the data. For example, it's becoming critical for organizations to back-up data and be able to recreate the state of the

data at any point in the past. These actions lead us to developments like the most recent ISO/ANSI SQL:2016 database standard or Amazon's Quantum Ledger Database (QLDB), which almost give you "data time travel" capabilities.

The financial services industry is placing greater emphasis on AI and machine learning governance, ensuring unbiased, non-discriminatory and fair AI. Regulators are following this trend (legislation is on the way).

Legacy technology and its replacement is the final driver of change. There's still an enormous amount of legacy tech out there, particularly in the financial services industry. Mainframes, for instance (not all are legacy, of course), often with tooling and architectures that haven't been updated for several years. Indeed, you need strict and coherent processes and strategies to work with legacy data, ensuring that QA is built into your processes. Any new fabrics and architectures must be engineered for future expansion.



Data strategy — the transformation journey

Let's look at the data strategies emerging from financial institutions and driving the evolution of financial data spaces. Some are generic; some are more specific to the financial industry — such as strategies for compliance with new data regulations.

Most organizations will follow a similar path toward data maturity, analytics and AI:

- Data management: Consolidation and curation
- Data democratization
- Data visualization: Self-service analytics
- Enterprise-wide AI, machine learning and decision support

During the initial data management phase, the organization should consolidate its data into one place. It's much cheaper to connect to and work with one location than multiple, diverse data sources. Teams can then curate data on an ongoing basis with automated tools.

In due course, there's a process of data democratization. Anyone in the organization who needs the data should be able to access and use it in their tool of choice, whether that's Excel, a visualization package or something else. Easy access also acts as an enabler for self-service data analytics and visualization with packages like Power BI, Cognos and Tableau.

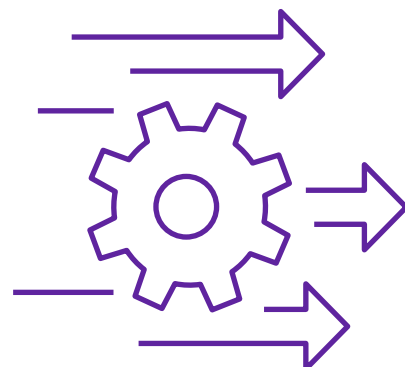
The next step is self-service (visual) analytics. Success depends on the quality of the data model to which your visualization package is attached. It's vital to get the right data model (or data environment) for a successful roll-out of self-service visual analytics. In other words, if users are to create dashboards, they'll expect the underlying data model to be correct, easily understood and "to do what it says on the tin." If this is not the case (e.g., the data is incorrect or poorly labelled), user trust will be lost and regaining lost trust is a long process.

Lastly, the data foundation is essential for implementing AI and machine learning, as is self-service visual analytics. Put simply, if an organization cannot create a reliable self-service model with underlying data models that are correct and substantial, then the chances of building an enterprise-wide AI and machine learning capability are slim.

Weaving your data fabric

A data fabric is a single environment consisting of a unified architecture with services or technologies running on top of that architecture. Stacks from many different providers now describe themselves as data fabrics. But the basic idea is to try and centralize things so that they're easier to govern and manage, and you replicate fewer unnecessary services.

The goal is to maximize data value, reduce the knowledge gap as much as possible and accelerate ongoing digital transformation.



Defining your **delivery approach**

How do we deliver this data transformation?

Organizations are outsourcing more and more of the data infrastructure and fabric. A decade ago, moving onto Azure, AWS or Google Cloud Platform was regarded as state-of-the-art innovation. Now, a cloud platform is just another service, and infrastructure and fabric can be operationalized, commoditized and outsourced, easily.

On the other hand, the amount of insight and intellectual property (IP) generated is also increasing. And firms are controlling this knowledge in-house, much tighter than in the past.

These are the two approaches to delivery. Organizations are keeping a tighter rein on data insights but are happy to outsource their data infrastructure.

Elements of **data management delivery**

The structure of the data management delivery has three key components:



IT delivery is an area with which most of us are pretty familiar. Increasingly, this is moving to agile models, combined with DevOps processes.

Data and model delivery mainly concerns your analytics model — an area that's maturing rapidly. The new issues are about how you manage and deliver data, AI and machine learning strategies. For example, regulators are increasing pressure on data versioning, so you can reproduce training results you had before

machine learning and audit any data changes. Also, you need to be able to explain how you arrived at your machine learning models and if you tested them for things like discrimination bias.

Regulatory and compliance delivery is also evolving rapidly. There's a swathe of new regulations coming out in 2022, including new EU regulations. So, it's vital to manage data privacy and security in a compliant manner, and be auditable.

Linking it all together

We can tie all this together with a target operating model that considers:



People

People need a broader range of skills. It's no longer enough just to say you're a DevOps expert. You need to train for regulatory issues in different jurisdictions and conditions. And to know that authentication and authorization requirements are correct, and to be acutely aware of jurisdictional issues and so on.



Processes

Processes have to be considered holistically and not in isolation. So, if you have a data analytics delivery project, you have to look beyond models, accuracy and confusion matrices to the various IT, regulatory and compliance aspects. As I'm sure you're aware, you cannot deliver an analytics project involving consumer data without a significant number of compliance checks and issues around data governance. You should factor these aspects into the overall project.



Technology

Technology transformation should be focused on increasing commoditization. Organizations need to concentrate on technologies that add the most business value and outsource everything else. Explore how you can outsource elements cheaply and effectively, and make sure the interface between in-house and outsourced components is seamless and secure.



Governance

Governance is critical, and the penalties for failure are increasing. Key areas to focus on include data quality assessments, cataloging, management, lineage and so on.

A journey of integration and change

In a nutshell, successful modern data management relies on integration across a much broader range of disciplines.

And the constant theme is that there's going to be change — and lots of it.

About **the author**



Paul Hewitt

Director and Practice Lead for AI, Machine Learning and Big Data, Luxoft

Paul leads the Data and Analytics practice for Banking and Capital Markets Consulting EMEA. His role spans all advisory and consultancy in data and analytics, and delivery of all projects from PoCs to large multiyear projects. He has worked with Luxoft for 3 years and has over 20 years' industry experience.

Take the next step toward data transformation

To find out more about the latest data management approaches, get in touch to continue the discussion.

[Contact Luxoft](#)

About Luxoft

Luxoft is the design, data and development arm of DXC Technology, providing bespoke, end-to-end technology solutions for mission-critical systems, products and services. We help create data-fueled organizations, solving complex operational, technological and strategic challenges. Our passion is building resilient businesses, while generating new business channels and revenue streams, exceptional user experiences and modernized operations at scale.

luxoft.com