Synechron

Assessing and Governing Data Quality in the Enterprise

The inventive use of data is one of the key battlegrounds between established enterprises and fintech challenger banks, with data quality a differentiator for success. But how can enterprises assess and improve data quality, and how can they use governance to drive the ongoing management of data quality? Data specialists from Synechron explore 'data quality' in the context of the financial services enterprise.

Niels
Daane

Consultant
UK

Why do you think there is increasing focus on data quality within financial services?

Financial services are an information business. Physical assets play no material role. All aspects of this industry are built around, rely on, or benefit from the ability to effectively exchange information. This effective exchange is only possible if actors trust the data they are handling. Key to this is putting data quality front and centre of a GBP 132 billion industry in the United Kingdom alone.

At an enterprise level, the case for high quality data should be obvious to anyone. High quality data allows for quick and effective decision making, accurate financial reporting, lower capital requirements and better customer services. In reality though, departments operate in silos, each maintaining its own data sets. Often, finance and risk functions use different definitions and employees in most firms still lose valuable time by handling incomplete or inaccurate data that requires them to manually reconcile data between sources. Such practices are particularly prevalent in large conglomerates, which have gone through mergers, acquisitions or corporate restructuring to expand their footprint across product lines, sectors and geographies.

Whilst an overwhelming majority of financial services providers continue struggling getting data right just to keep their business running, one recent development puts even more emphasis on the ability to trust data. With challenger banks and big tech entering the financial services domain, the pressure on incumbents to remain relevant to their customers continue increasing. But with up to 90% of their staff fully engaged with keeping the business running, they are inherently poorly positioned to gain ground in the changing landscape.

In response to the emergence of these new players, financial services providers are heavily investing in building out their Innovation and Analytics capabilities. These teams consist of strategists, design thinkers, data scientists and developers, often located on separate floors or even residing in a building off the main premises. Although independent from business as usual and its bureaucracy, these teams are still very much dependent on the data that is collected, stored and exchanged as part of the firm's core processes.

Although machine learning models are scalable, they rely on a pipeline of accurate test data. This is highly relevant especially where these innovation teams have adopted DevOps and Agile methodologies for iterative model development. With poor data quality cited as one of the biggest bottlenecks for machine learning adoption, unavailability of high quality test data will seriously impede an incumbent's abilities to pivot and remain relevant to their customers.

At present, traditional players still have the advantage of long-established client relationships due to their role as a trusted expert advisor. Whilst their rich customer databases could provide a competitive edge over new entrants, this data is only useful if it can be trusted. If this data is not of sufficient quality, innovation will remain limited to local pockets of data scientists creating their own metrics and dashboards, which cannot be meaningfully scaled up. The required remediation of data issues, however, often remains limited to tactical solutions. These become permanent as attention and funding flows to other, more exciting initiatives with higher priorities. A pragmatic approach to data quality management is possible and can coexist with business as usual and strategic initiatives.

In your opinion, how should responsibility for data quality be assigned within an organization?

Everyone who touches data bears responsibility for its quality. Uninformed decision making or inadequate handling of anyone who interacts with data can result in poor quality data. Thus, improving data should be a continuous effort, focused on the entire data life cycle. This will span across separate departments.

However, because data is everywhere and is constantly on the move, it is difficult to formalize responsibilities. Even so, appointing owners for all data elements will not be sufficient for improving data quality in your organization. Rather, it requires senior buy-in, adequate controls and a cultural shift towards protecting data quality.

At an enterprise level, this responsibility is best taken by a dedicated program or department tasked with improving data quality throughout the organization (e.g., a Data Management Office). It should be noted that we consciously don't use the word 'project' as that would imply that solving data quality issues can be completed in a predefined time frame. Instead, it is more likely to be a continuous effort, which could eventually become an integrated part of the organization.

The data quality program itself should not take ownership for all data. At most, it can take responsibility for reference data, as that is used across business lines. Instead, it should set out and establish an ownership model for the different types of data that are recognized within a firm. For instance, it can decide that ownership for client and transactional data sits with respective business lines where this data originates, with the people who know what good looks like. These teams will be best positioned to single out errors and instruct due data remediation efforts.

If done properly, assigning responsibility of data can inspire a cultural shift on its own accord. Combined with data lineage throughout business lines, it will finally be clear to employees with whom they can address some of the persistent data quality challenges they face. This cultural adoption can be further solidified by formalising responsibilities in job descriptions, and making them part of individual and department-level appraisals. Adoption of this 'culture for quality' should be championed by senior executives, so that it transcends across all departments and layers of the organization.

What steps do you recommend to drive the improvement of data quality within an organization? What areas should be in focus?

Focusing on ad hoc incident resolution and data corrections, firms will fail to identify and address recurring data quality issues in a structural manner. Wrong data is an outcome of interactions between people, processes and systems and it can be complex to solve these issues. Recognizing this is the first step towards meaningfully improving data.

When setting out to improve data quality, firms should ensure they come up with a clear data strategy that aligns data handling with the organizational goals. This strategy will inform a series of subsequent steps an organization should take to improve their data quality.

These steps should include:

1. **Creating** a clear data strategy
2. **Scoping** the focus on critical data elements
3. **Defining** what good data looks like
4. **Cleaning** data and establishing data lineage
5. **Embedding** data quality controls
6. **Defining** clear roles and responsibilities

To ensure that this program is focused and pragmatic, an initial data quality assessment can be carried out to capture what is already in place. On top of that, it is important put focus on the most critical data elements and to involve senior executives from the start. It will already be difficult to improve these key data elements, as thinking and speaking about data will advance ideas about what data should look like.

As data quality efforts mature, a next level of sophistication in control and governance can be adopted. The first, foundational level, involves the acceptance and use of standard measures and reconciliations to ensure data quality. In a more advanced stage, an enterprise can apply anomaly detection using statistical analysis at key data points. The most sophisticated level involves machine learning to analyze patterns, to pre-empt emerging data quality issues, and assist remediation by suggesting correct values for suspected errors. Each level will allow staff to spend more time on value-adding activities, such as helping their firms to stay ahead of their competition and remain compelling to their customer base.

Tim
Jennings

Associate Partner
UK

When you are assessing an organization for data quality, what are you looking for and how do you measure it?

We are all familiar with the standard dimensions of data quality that explore elements of data precision, accuracy and consistency. These are useful measures for specific data items or feeds, but it takes a significant amount of pre-work to use these to assess whether an organization has intrinsic data problems. So where can you look instead to get a top down perspective? There are a few useful areas to start.

The first is in the data architecture of an organization:

- Is there evidence of duplicated mastering, where the same data items are being generated from different systems?

- Is there evidence of data being passed between non-mastering systems and so subverting the golden source?

- Is the timing of published updates synchronised with the usage expectations?

- Are there significantly different data models in use to describe the same data?

Where these architectural problems exist, a useful way to test the impact is by looking at the reconciliations that are taking place. These develop organically at the data weak spots as Operations or IT teams try to stem the tide of data quality issues. So, how many reconciliations are taking place, at what cost, and with what impact? Where are the rec breaks occurring? What problems are they finding? How many corrections are being manually applied on the data systems? In a similar way, the number of corrections/updates can be a business level indicator of data entry or data capture problems that ultimately undermine data quality.

Another way to assess the quality of data is to look at the quality of the reporting. It is not always easy to see whether data is accurate at the time of publication but there are often processes that issue corrections or retrospective updates, driven from business quality control checks. These can be another useful indicator of problem areas in the data estate.

How do you convert this into ongoing KPIs for data quality assessment?

It is reasonably easy to see the metrics that can be derived from the top-down approaches to data quality. Data masters, rec breaks, manual corrections to data, or updates to published information are all useful ways to measure quality.

If you are instrumenting your organization for data quality, now is also the time to start including the typical 'ground-up' measures and compiling these into KPIs. Individual data producing systems can be tested for the Precision metrics around completeness, uniqueness, specificity and conformity. These can normally be automated to produce regular comparable reports that allow you to track data quality over time. The Accuracy dimensions of staleness and correctness may be less easy to automate (owner attestation may be needed instead), but again these can be compiled on a system level and combined into a set of KPIs. All of the metrics listed here are the responsibility of the producing system to manage from a quality perspective.

For consuming systems, the key measures are: consistency timeliness, coverage and integrity. Again, testing of some elements can be automated or architected for.

**There are a variety of tools available in the data quality space offering features from data profiling to crowd-sourcing trust scores.
Are there key features that you would recommend which offer a dependable path to improving data quality?**

Of course, any tooling can only be as good as the process that it feeds into – tooling alone will not fix your data quality problems and will not necessarily be able to identify where your data is not fit for purpose for your processes. The tooling in this space provides a variety of different views of the data, and can certainly test some of the precision dimensions mentioned earlier, but it offers no silver bullet. In my view, the key to delivering dependable quality lies in:

- **Identifying** the correct quality measures for your business process/es

- **Automating** the quality assessment testing and collection of metrics for that data

- **Establishing** a proactive process that seeks to improve the measure scores

This will often require a variety of approaches, from data profiling to custom feeds analysis, etc., and no one vendor tool is guaranteed to have everything you need.

The use of crowd-sourced data scores serves a rather different niche than data quality metrics – it is more an indicator a 'usefulness' of self-service datasets although the consumers may not understand the rationale the proposer used when scoring the dataset. It certainly can be useful, and it is likely that a dataset with a larger number of good scores will be a reasonable place to explore if you are looking for usable data. This does not guarantee that the data is from golden sources, that it is complete, or complies with expected standards, etc.

About our SMEs:

Niels Daane
Consultant
UK

Niels works within our Data Governance practice at Synechron in London and focuses on data management. He works with colleagues across our Data Science & Analytics teams, and is committed to further developing practical applications for data which plays a central role in creating understanding between data scientists and business people within the Financial Services industry. He joined Synechron in early 2018 after achieving his Master's Degree in Organizational Studies in the Netherlands.

Tim Jennings
Associate Partner
UK

Tim works in Synechron's FinTech and Digital Enterprise practices, focusing on setting and executing strategic business and technology change for Financial Services firms. He has 20+ years' experience in Financial Services, with practical knowledge of business transformation, developing strategic IT and data architecture, and leading adoption for Front Office, Operations and Control functions.

To know more about Niels

https://www.linkedin.com/in/nielsdaane/

To know more about Tim

https://www.linkedin.com/in/tim-jennings-b99b201/

Synechron