Beyond Firefighting

A Leaders Guide to Proactive Data Quality Management

An Experian Whitepaper

Experian™
Data Quality

Table of contents

About Author

Dylan Jones has 20 years' experience of delivering large and complex data quality and data migration initiatives across a wide variety of corporate sectors. In recent years he has devoted his time to promoting industry best practices via the community websites Data Quality Pro and Data Migration Pro, educating and supporting hundreds of thousands of professionals across the globe.

Author
Dylan Jones

Editor,
Data Quality Pro/Data Migration Pro

"I don't really understand why you're here, we don't have any data quality issues".

These were the words of a utilities data manager upon my arrival to assess their data as part of a corporate transformation scheme.

It's common for some managers to react in this way because in their eyes there are no issues. The business goes about its work, there are no embarrassing stories of poor data quality hitting the headlines and everything seems to be operating as expected.

But scratch the surface and things are not what they seem.

In my utilities example above we started to discover some startling findings within the first few hours of assessment.

20% of equipment had no manufacturer codes assigned. 30% of buildings could not be identified via postal address. An accuracy check uncovered 35% of equipment was not located in the location specified in its engineering plan.

The list went on and when we presented the findings to the data manager a stream of excuses came flooding back.

- "We don't need to know the address of buildings, our engineers know just from the first line of data"
- "Those manufacturer codes are meaningless, we don't use them any more"
- "Those guys in engineering are always sending us out of date installation plans, we're the ones who know where to put the plant equipment where it makes most sense"

In the eyes of the data manager there were no issues because the field workers simply accepted the flaws and "worked around them".

This "working around" data quality issues invariably means "scrapping and reworking" to get the job done.

- The engineering plan doesn't resemble reality? Scrap it and rework a "more accurate" local copy.
- Health and safety want to send an announcement to each building address on file? Scrap the original plan of a simple mailshot and rework by laboriously cross-checking building details against other reference data.
- A manufacturer wants a recall on a particular product? Scrap the idea of creating a consolidated national report of all equipment types and location, rework by assigning a staff member to contact each local engineering team over several weeks to track down missing products.

You can of course substitute retail, pharmaceutical, healthcare, banking, insurance, government and any number of sectors but you'll see the same result time and time again.

Reactive, localised data quality workarounds can cause huge impacts to the organisation because they transfer valuable resources away from value-added work.

The end result is a range of damaging impacts including:

- Reduced customer satisfaction and increased customer churn
- Inability to outperform competitors and win new business
- Operational inefficiencies and wasted budgets
- Reduced confidence in strategic decision making
- Regulatory and internal compliance non-conformance

The dilemma facing companies of all sizes is how to break out of this endless cycle of data quality firefighting and prevent these impacts occurring.

Armed only with limited knowledge and experience of data quality management, how should companies progress to a more proactive approach? How can they start to prevent defects at source and provide controls to ensure any defects that do slip through are promptly dealt with?

The answer is to develop a series of key data quality management capabilities that have been repeatedly proven to foster a more proactive approach to data quality management.

Let's explore some of these in the next section.

2.1 Where should we start?

This is typically the first challenge facing the leader who wants to move their organisation towards data quality maturity.

There are typically no shortage of data-dependent initiatives taking place at any one time and certainly no shortage of defective data to tackle but where should those tentative first steps towards proactive data quality be taken?

There are no hard and fast rules but when getting started it helps to align your data quality efforts with a strategic initiative that has similar goals.

For example, a lot of data quality management initiatives are kick-started by supporting Lean, Six Sigma or other quality improvement initiatives.

Other areas of focus may be Risk, Compliance or Governance initiatives.

Perhaps your executive team has highlighted comprehensive cost-saving or operational improvement strategies as part of a corporate scheme. Are there major data migration or data integration initiatives that will almost certainly take far longer without proactive data quality support from the outset to ensure that future business functions are not impacted by poor quality legacy data?

Proactive data quality management can support all these initiatives and many more of course so, where possible, collaborate with an existing programme that has executive backing. Demonstrate how the end result will be more successful and achieved faster through data quality management.

You will almost certainly need to provide a pilot scheme that demonstrates the challenges and delays they will face without the support of your team.

In order to deliver this pilot and build on any early successes you will need to adopt some core capabilities of data quality management.

'You will almost certainly need to provide a pilot scheme that demonstrates the challenges and delays they will face without the support of your team.'

2.2 What are the core capabilities for proactive data quality management

Data Lineage Management

To truly measure the quality of your data you need to fully understand its provenance.

Do you know the various routes and pathways your critical data takes from creation to consumption? Do you know who owns the data and is responsible for its stewardship as it passes from system to system?

Proactive data quality management requires you to track the most critical data in your organisation so that any defects can be resolved at source. To prevent reactive and localised you need to look beyond the boundaries of your own department or system silo to where your information originates. Work with these groups to fully document the lineage of your data so that you can not only resolve defects more efficiently but begin the reengineering of these information chains so that data is validated and prevented at source.

Data discovery products can provide considerable support here. To map the data pathways and relationships manually can take many months, even years. By adopting data discovery technology you can automatically discover relationships and data lineages, saving your team months of effort. Once these pathways are defined you can then instantly monitor and report on data quality levels from creation and onwards through the entire organisation.

Business Glossary and Data Quality Rules

To measure data quality you need to fully understand the definition of data and the rules it must obey.

For example, consider the attribute "CUSTOMER_ACQUISITION_DATE" - what does this really mean? The date the customer became a record in a database? The date they purchased a product? What rules would we expect for this value?

To create high quality data you need high quality definitions and comprehensive data quality rules. These should not be archived away in some ageing Excel spreadsheet but should form the very core of your proactive data quality initiative.

By embedding these rules and definitions directly in your data quality technology platform you can instantly measure the conformance of your entire data landscape. Adopting a data quality technology platform that allows this central management capability is critical in keeping your costs down and scaling your reach.

By creating a Business Glossary of your most critical data you can instantly verify where breaches are taking place across the organisation. By combining with Data Lineage Management you can then trace downstream impacts and notify dependent information consumers who may be aware of the defects you're observing.

Data Quality Management Framework

A data quality management framework provides an organisational blueprint for how all the different components and capabilities required for proactive data quality management interact.

Without a framework it becomes difficult for the business to gauge the bigger picture and understand where their involvement lies.

Data quality management frameworks help you benchmark your own approach and identify gaps in your own planning or approach. This is to be expected in the early days but you now have a roadmap for which components are missing and can be developed organically as your impact on the organisation develops.

Which Data Quality Management Framework should be adopted?

There are no hard and fast rules here, there are many available, however here are some considerations:

- Is the framework proven? Can others testify to its value and impact on their organisation?
- Does it cater for different styles of data quality initiative? Can it address the full spectrum of data quality requirements?
- Does the framework cater for organisations at various levels of maturity? Does it provide a transition from one stage of maturity to the next?
- Is the framework accessible? Does the framework come with education and training programmes so that your team will be able to adopt the techniques quickly?
- Can the framework be implemented using data quality technology and other tools so that it becomes embedded within the organisation?

Above all, your data quality framework needs to be easily understood and remember that you can extend the framework as your organisation matures its attitude to data quality.

Data Quality Technology Platform

Perhaps the biggest challenge you face in maturing proactive data quality management is scalability.

You may start with a small team or perhaps you are the "team". How can you hope to scale your "sphere of influence" and impact real change within the organisation with such limited resources?

The answer lies in two main areas:

1. Create a distributed data quality capability
2. Appropriate Data Quality Technology

A single data quality team can not raise quality levels across the entire organisation, it will simply take too long and you will be swamped by the scope of the task ahead. Instead you need to develop a federated approach where you nurture data quality skills, ownership, stewardship and standards across the organisation so that a single data quality team doesn't become a bottleneck. In effect you're "baking data quality into the organisation" so that data quality processes and procedures become business-as-usual.

There is a problem however, a distributed capability takes time and you need to scale right now. Data quality technology can help you here and must form part of your "data quality war chest".

The right data quality technology will afford you the ability to scale a proactive approach to data quality because you can instantly automate literally thousands of activities that would normally be carried out through manual effort.

What's more, the right data quality technology can perform across the full spectrum of data quality challenges, from validating and eliminating defects at source, through to continuous monitoring and defect resolution management downstream.

Take the typical data migration initiative as a prime example of how data quality technology can help a project team deliver a better and faster end result. Traditionally, some basic audits or checks would have been carried out during the data migration using whichever scripting tools were available. As the data is migrated across to the target system a lengthy process of manually fixing any failed records would ensue. Once the data migration was completed another team of testers would have to painstakingly solve data defects in the target system that users would begin to identify.

All of these problems are eliminated with a robust data quality strategy backed by the right technology. The performance benefits of adopting the right technology on a data migration can be staggering. Tasks that would once take weeks, even months, can literally be completed in hours if you have the right data quality technology at your disposal.

Of course data migration is just one initiative. Imagine all the other data-dependent projects, not to mention everyday business-as-usual operations, where data quality technology can proactively resolve data quality issues.

Reference Data Strategy

Reference data is information that is sourced from outside your organisation that can help you gain deeper insight into your own data whilst acting as a trusted source to compare against your own data. As such, it can provide a powerful tool in your quest to proactively improve data quality and demonstrate real value to the business.

In my earlier utilities example the issue of incomplete address data was presented to the business. The data manager hit back at our data quality assessment stating that they could see no value in having good quality address data as the engineers "knew their own patch".

However, we persevered and improved the address data for several regions of the country. We were then able to overlay their entire engineering plant on a geographical display of the UK. The lightbulbs started to switch on.

"We can schedule engineer visits based on nearest engineer!"

"We can plan equipment swapouts by geographic region"

"We can see which locations are failing to adopt appropriate naming conventions"

By cleansing the address data with an accurate and trusted data source we were first able to dramatically improve the quality of data found in the engineering data. Overnight

their information became accurate, complete, formatted and useful. However, we were able to go a step further. We could now link the address data with geocode information to create latitude and longitude coordinates for each building. This added dimensions opened up whole new operational solutions that the management team had never even considered.

Introducing reference data does this to a business. It creates innovation and competitive advantage. These are precisely what the business wants to see from your team.

Proactive data quality management is not just about improving processes, it's about creating unforeseen opportunities that result from the often dramatic improvements in data quality that your team can produce.

Always keep your eyes open for opportunities to introduce new data to existing data sets. The benefits will be attributed to your data quality initiative and when starting out this is exactly the kind of good news story you want to promote.

Ownership Framework

One of the biggest causes of reactive and uncoordinated data quality initiatives is a lack of data ownership.

Data needs an owner. Someone has to step up and be accountable for the quality of information that is driving your operational processes and strategic goals.

Many companies believe that data ownership lies with a function of the IT team. This is a popular misconception and one of the primary reasons that a big divide often results between the business and IT. The business often feels like IT is not resolving their issues quickly enough and IT feel like they continuously "cop the flak" for the poor quality of data.

Whilst IT can obviously play a role in developing the right systems to store and process data they should not be responsible for ownership of data content. This must fall to the business because it is they who truly understand the data. It is the business who create and maintain the information so therefore they should ultimately become the custodians.

This is not an easy pill to swallow for many, particularly those in senior management who often feel that they or their teams are already swamped with a heavy workload.

In a sense these are well founded concerns. Data domain experts make ideal data custodians and these workers are always going to be in-demand because of the knowledge they've amassed.

So how can you create a team of data owners and custodians around the organisation from an already busy workforce?

Once again data quality technology can help.

Modern data quality technology enables data owners to instantly assess the quality of information across their domains and subject areas. They can track performance over time and identify areas of concern that require their involvement. Once they have defined the standards and rules that data in their domain must conform to these "data quality specifications" can be managed in data quality tools so that data owners can focus far more on setting the vision and strategy for their data instead of getting bogged down in operational tasks.

There is a natural hierarchy of ownership. At the top you may have a CDO (Chief Data Officer) who sets the vision and goals for data owners across the company. Moving down the hierarchy you may require Subject Area Data Owners, people who are

'By cleansing the address data with an accurate and trusted data source we were first able to dramatically improve the quality of data found in the engineering data.'

responsible for perhaps Customer Data across the organisation. You may even have multiple functions involved in this role, perhaps legal and compliance, content and quality, design and development.

As you move down the hierarchy further you may have a stewardship framework, those domain experts who actively monitor and report on data quality across information chains or primary business functions.

Whatever your ownership framework, try and ensure that you are supporting these key resources with the right level of technology. It will ensure that their focus is on continuous improvement instead of endless administrative chores.

Root-Cause Analysis and Prevention Strategy

Finally, we come to the most important element of a proactive data quality management approach - root-cause prevention.

To move from a reactive, fire-fighting mentality, requires a major re-think in how your organisation traditionally tackles data defects.

A classic example was found in an energy client that required a member of staff to clean up inbound feeds of data from its installation contractors. Every week this individual would waste 2-3 days processing, correcting and requesting a revised copy of the inbound data until the information was entered into the customer management system.

Invariably defects slipped through because the poor soul was trying to fix the data by hand. The impacts to the business were considerable. Longer lead times for new site installations, billing and invoicing delays, inaccurate field force manifests to name but a few.

To resolve this issue we taught the worker to become self-sufficient in data quality techniques.

They were given basic training in data profiling and data quality rule creation.

They were taught how to create data lineage pathways that modelled the flow of the data. They were taught how to integrate external reference data so that they could cross-check the inbound data and extend it with additional information about the location and customer details. They were taught how to document the data owners, both upstream and downstream, so that alerts could be sent promptly when issues were found and creative solutions discovered.

Most importantly, they were taught how to implement root-cause analysis techniques. They were taught techniques such as “5-Why’s” so that they could consult with contractors as to why certain defects kept re-appearing and what could be done to prevent them.

After several weeks they were able, with support from the central data quality team, to create a persistent “data quality firewall” that continuously monitored inbound data for quality breaches, instantly reporting back to the feed owners the precise location and nature of the problem.

This root-cause prevention strategy had profound benefits on the organisation and not least on the worker in question. They became more productive, proficient in data quality and a leading advocate for quality improvement across the department.

‘The impacts to the business were considerable. Longer lead times for new site installations, billing and invoicing delays, inaccurate field force manifests to name but a few.’

3. In summary

The data quality firefighting dilemma impacts all organisations at some point. All organisations struggle to break out of the endless cycle of manually fixing defective data.

However, the organisations that face the challenge head-on and implement some of the capabilities discussed in this paper go on to reap the rewards. They deliver better, faster, cheaper solutions that delight their customers.

What is the main difference between organisations that stagnate and get caught in the endless firefighting trap and those that take the leap of faith and adopt a proactive approach to data quality?

Leadership.

Someone has to recognise that data is one of the most vital assets an organisation possesses. It holds the key to competitive advantage and world-class performance but it requires strong leadership to ensure it is valued and protected.

As data volumes rise exponentially and the dependence on high quality information assets increases the old tactics simply won't scale or deliver the kind of business your organisation demands.

Your role as a leader is to create a roadmap beyond the reactive world your data currently inhabits to a more proactive, preventative type of data quality management.

Hopefully this White Paper has pointed the way.

About Experian Data Quality

Experian Data Quality has built up exceptional market coverage assisting customers with their unique data quality challenges.

We provide a comprehensive toolkit for data quality projects combining our market leading software with a vast scope of reference data assets and services. Our mission is to put our customers in a position to make the right decisions from accurate and reliable data. The size and scope of data management projects varies considerably but the common factor in all ventures is unlocking operational efficiency and improving customer engagement. We see the potential of data. Whether it's in enabling ambulances to be sent to the exact location of an emergency or attributing charitable donations to the people who need it the most - data accuracy makes all the difference to service provision.

For more information please contact us:
T (03) 8699 0100 | E info@au.experian.com | W www.qas-experian.com.au