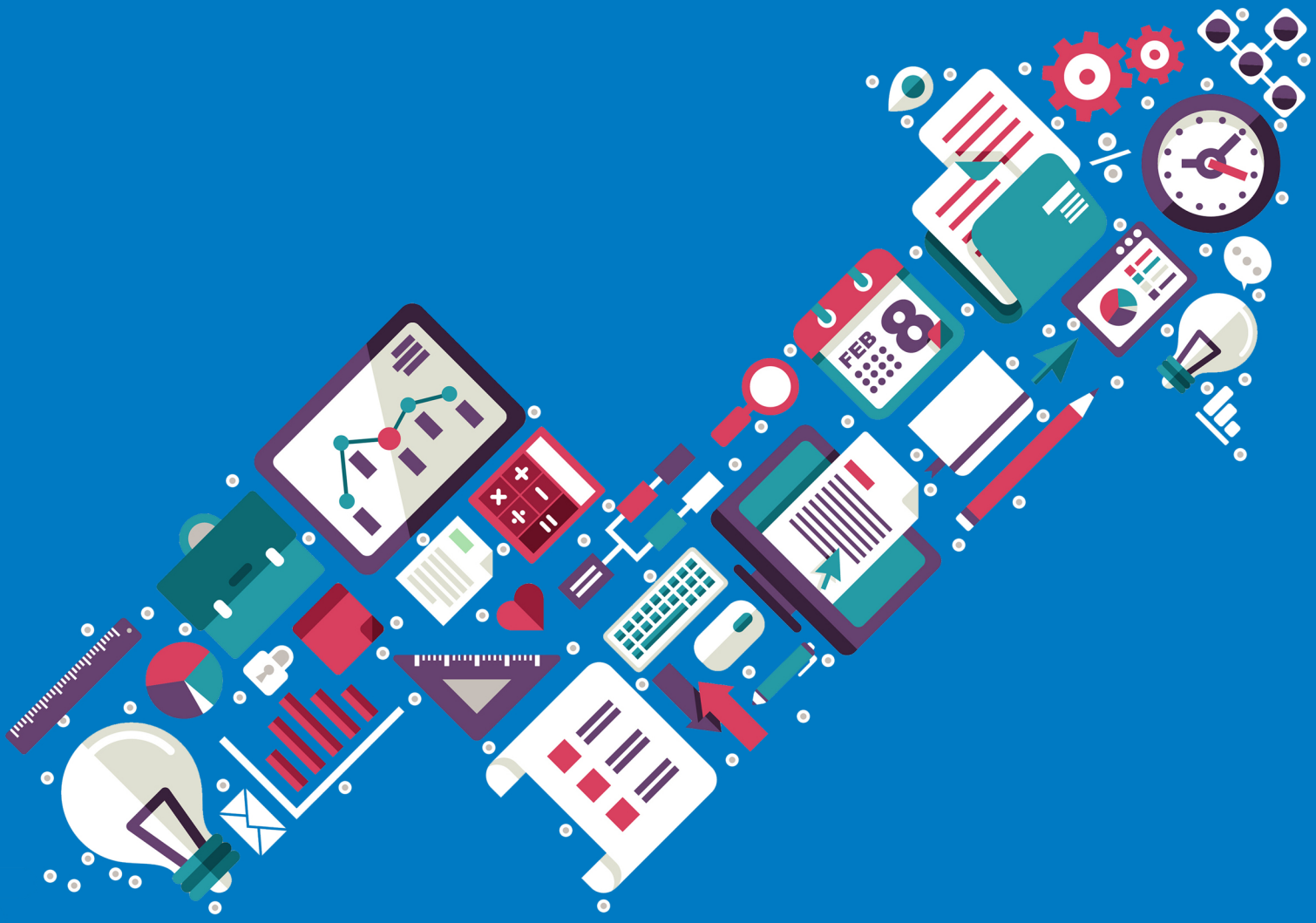
# How to Create a Comprehensive Data Quality Management Strategy

by Dylan Jones

Editor of Data Quality Pro




Experian™
Data Quality

August 2014

# Table of contents

# Section 1: Introduction

Most organisations today recognise the vital importance of data in fulfilling their strategy objectives. In the recent 'Global Data Quality Research 2014' Experian Data Quality Market survey, approximately 99% of companies had some kind of data quality strategy for at least some part of their overall data landscape.

This growth in data quality adoption is a welcomed trend but digging deeper into the findings we see that despite this increased focus on data quality in modern organisations, the performance of these data quality efforts are mixed.

Many organisations are dissatisfied with their data quality management approach and, even where specialist data quality software is being used, 55% are not happy with the results.

The reality is that many companies are still using outdated methods of data quality management that are often reactive in nature and can result in far longer lead times to actually resolve any issues found. Tools such as Excel and SQL often complement manual analysis and resolution tactics. All of this means that data consumers, and indeed customers continue to suffer the effects of poor quality data for much longer periods than they should.

Compounded by the lack of a correct data quality management approach, many companies are omitting key areas of their data landscape. One in three large British organisations do not include all of their data in a data quality strategy and this creates pockets of poor quality data that can impact any services that depend on the underlying information.

The key areas of concern where organisations are clearly keen to improve relate to the Detection, Analysis and Resolution of data defects. These activities form the backbone of any data quality strategy and a flawed approach in any one area can seriously undermine the overall data quality management initiative.

The challenge of course is to help companies that are dissatisfied with their data quality results shape a new direction by adopting better approaches for data quality management.

The aim of this report is to help organisations create an improved data quality management strategy by drawing on insights gathered during hundreds of practitioner interviews with experts who have successfully matured their organisations' approach to data quality management.

The report will focus on the key areas of data defect Detection, Analysis and Resolution but will also highlight some other initiatives that will benefit a more holistic data quality strategy.

## Section 2: Why Do Organisations Struggle with Data Quality Management?

One of the biggest problems that many organisations overlook is the need to manage data quality holistically. Far too many people take a localised view of data defect management and simply apply quick fixes that suit their own functional or departmental needs. The common trait amongst all organisations who take a mature approach to data quality is the need to understand the much broader chain of events that can lead to poor quality data.

A classic example of this was the case of errors found in a utilities organisation that had several systems used to record the placement and operations of power equipment.

The organisation found many instances of location and equipment records having poor quality data. Some of the issues included:

- Partial building address data
- Missing floor or bay data
- Duplicate equipment sharing the same floor placement data
- Entries for locations that no longer existed

In some cases they found as many as 35-40% of records having some form of data quality issue and yet the organisation was still functioning and delivering services to their customers.

How is this possible? How can organisations have this level of poor quality and yet still deliver profits and satisfied customers?

The answer lies largely in reactive processes. If an engineer arrives at a site to find duplicate equipment in the same location they simply create yet another record and get on with their job of installation.

If a site address lacks information then the engineer checks other systems or uses their own local knowledge to work around the problem.

When a customer reports that their communications have gone down because equipment was incorrectly moved, the engineering team is quickly dispatched to resolve the issue.

Organisations deal with poor data quality by reacting to conditions as they arise. This is clearly a costly and short-term strategy because the root-causes of these issues may still persist.

It is a myth that there are single points of failure with data quality issues. Quick fixes may paper over the cracks but there are often much broader problems that stem from failures with policies, procedures, technology and training (to name but a few).

In the earlier utilities example it was found that many of the defects actually stemmed from a variety of causes including:

- Lack of appropriate training and awareness of the importance of data quality
- Conflicting rewards structure that had a bias towards getting the job done quickly at the expense of correct records capture
- Insufficient validation and control logic at the point of data entry
- Integration and migration issues between operational systems
- Lack of routine auditing and accuracy controls

You can therefore see there are typically no single points of failure with data. You need to look at the broader picture and understand the entire data lifecycle before making decisions.

The key takeaway from this section is not to think of data quality defects as some kind of isolated incident but potential indicators of much wider issues that often require people, process and technology intervention to resolve.

In some cases they found as many as 35-40% of records having some form of data quality issue and yet the organisation was still functioning and delivering services to their customers.

## Let me provide an example of one organisation I interviewed that took a more holistic and mature approach to data quality management so you can see what the future holds for your organisation if they introduce some shifts in data quality strategy.

This organisation provided complex engineering solutions for large corporations but were effectively like every other company on the planet - they used information to deliver customer services.

The service they provided would take several weeks to complete and any defect in the process, particularly with the data, would cost them significant drops in profit due to unplanned delays.

This company adopted a holistic approach to data quality management. They had a monitoring process that observed the performance of every aspect of each customer assignment and senior management had a 360 degree view of data quality levels across the entire business unit.

To enable this the data quality team had mapped out all of the business functions that were required to deliver the customer solution. They then identified all of the information sources that supported those functions and created a library of data quality policies, standards and rules that could be monitored hundreds of times a day.

Combined with these data quality metrics were other operational metrics such as process lead time, cost per unit, items in spare and reported faults.

This combination of data quality metrics and business performance insight enabled management to observe the impact of having high quality information supporting their business.

But the benefits didn't end there.

Because this organisation had created an entire information chain of the customer delivery process they could instantly pinpoint data quality defects and track them back to their source. They could see where other systems were dependent on this information and which stakeholders needed to be identified. If software changes introduced data defects into the system then these could be quickly identified because an holistic early warning system was in place.

If this sounds like some kind of futuristic nirvana then please bear in mind that all of this was achieved over a decade ago using technology that would be considered outdated when compared to the latest data quality management technology.

What's more, modern data quality technology allows this data quality landscape to be implemented in a fraction of the time of legacy solutions at considerably less expense than a manual, reactive approach.

## Focus Areas

In the recent Experian Data Quality Research it was discovered that many organisations are struggling to implement adequate Detection, Analysis and Resolution processes as part of their data quality strategy so we will focus on what is required to improve these activities.

We will also explore some aligning capabilities that will help you deliver a more comprehensive data quality management strategy.

### Phase 1 : Business Function and Process Discovery

It is easy to think that data quality management should start with a focus on data but we need to step back a moment and consider exactly what functions our data needs to support. Omitting this step leads us to put on 'data blinkers' and only focus on the data we think is important.

By understanding the many functions and processes that our data has to enable we start to take a much more business-oriented view of data quality. This helps us later because it ensures that when we do find defects we can instantly see how that relates to the core business functions of the organisation.

To discover Business Functions is quite a straightforward task and is typically carried out in a workshop format. You speak with senior stakeholders about the vital operations they are responsible for. From there you map the high level functions of the organisation. These could be things such as service fulfilment, billing operations, order handling, recruitment - whatever is relevant to your organisation.

Underpinning these high level functions will be a series of other functions that can only be discovered by talking to actual business users, again in a workshop format.

Following these discussions you arrive at 'Candidate Business Functions' which are the building blocks of how your organisation does business. What you will also discover during these workshops is precisely which data stores are used to allow these functions to successfully complete.

As you start to analyse these business functions further you can discover the business processes and sequence of events that underpin each function. You don't need to be a business process expert to do this, simply creating a process chain on a wall with Post-It notes is a great starting point.

What this activity will enable is a much clearer view of:

- The most important functions and processes your data supports
- The stakeholders and subject matter experts who understand and govern these processes
- The chain of events that align to the underlying data

By understanding the many functions and processes that our data has to enable we start to take a much more business-oriented view of data quality.

For more information on how to conduct a Business Function Discovery workshop visit: http://bit.ly/business-function-discovery

## Phase 2 : Information Chain Discovery

You should now have a much clearer view of how your functions, processes and data stores align. Now we need to connect the dots with the data and see how your information chains support these core processes that drive revenues and profit for the organisation.

To do this you need to draw on the information gained in the earlier workshop but ideally supplement this information with a data discovery capability.

For example, the workshop will tell you that the order handling representatives you spoke to received customer orders via the telephone or mail order but there may be other customer request information flowing into the order fulfilment function.

Data discovery functionality in modern data quality tool allows you to supplement this local business knowledge by automatically discovering relationships in the data without any prior knowledge of how that data is connected.

You simply point the tool at your data landscape and it will find the links between the systems in seconds. This is a powerful capability to have when understanding the construction of your information chains and can save many weeks of effort compared to manual approaches.

It also helps to validate prior assumptions about how systems are connected as well as give you immediate insight into the quality of the information chains.

Can you see how effective this activity becomes when trying to discover the root-cause of issues? By automatically discovering these information pathways in your organisation you can track the lineage of data in seconds as opposed to carrying out laborious manual analysis using unsuitable tools such as SQL and Excel. Indeed, this delay in finding the source of defects was a major cause of concern amongst interview subjects in the recent Experian Data Quality Market Research Report.

---

For more information on the process involved in creating an Information Chain the following publication by David Loshin provides an excellent chapter on the topic: ‘Enterprise Knowledge Management: The Data Quality Approach’

http://bit.ly/loshin-chains

---

## Phase 3 : Data Quality Rules Discovery and Management

The goal so far has been to understand what functions the business performs so that we can prioritise and tailor our data quality strategy accordingly. Hopefully you now understand that taking a business-focused approach means that any subsequent data quality work we do will maximise the benefit for business users and indeed customers.

Once we have identified the core information chains that drive an area of the business we will go one step further and understand what rules our data must conform to. This is a critical step as these rules dictate what types of quality we want our data to be measured against and also what thresholds and quality levels we will monitor in future.

To do this we need to draw again on the support of our business users because they possess knowledge of the main operating procedures and local policies that dictate whether the data is fit for purpose.

However, using a modern data quality tool that possesses data profiling and data quality rule discovery functionality means we can considerably reduce the time taken to discover our data quality rules as well as creating a much richer library for future use.

We already know which data stores are in scope for observation so by profiling the data we can discover huge amounts of rules for categories such as:

- **Attribute Domain Rules**: What values should be expected? Should they be empty or populated? What is the expected format? How do the rules vary for different subsets or conditions of data?
- **Integrity Rules**: What attributes should be unique? Which data elements should be related? What type of relations are expected?
- **Event, History and State Rules**: What sequence of events is the data expected to show? For example, are there any specific states that equipment or customers must be in to be correct?
- **Dependency/Consistency Rules**: How does one data attribute influence the value of another? What is the expected value of one attribute given a certain combination of values in other attributes?

These rules form the bedrock of your data quality strategy. Organisations often only focus on a large quantity of basic data quality rules at the expense of understanding a smaller number of more complex rules. It is often the more complex rules, perhaps spanning many systems, that hold the key to data quality success. When these are well managed and continuously improved you will start to see immediate benefits to the business.

It cannot be stressed enough that trying to discover and manage these rules without a data quality tool is virtually impossible. The length of time taken to discover them is more than enough justification to warrant the introduction of a data quality tool but where tools really come into their own is the ongoing management, maintenance and monitoring of these rules. You simply cannot achieve this cost-effectively with manual or 'home-grown' solutions.

Once your rules are discovered they can be stored in a library and managed via your data quality tool and reused across the organisation. For example, if you create a rule to measure certain identifiers that are common across the whole organisation then this rule can be shared and managed widely, providing immediate long-term value to the organisation.

The benefit of storing rules in a modern data quality tool also means that you create an asset that the business can use as opposed to some in-house technical scripting solution that only selective individuals in the IT circle can manage. Data quality management is a business discipline and modern data quality tools enable a release of that control from IT back into the business.

## Phase 4 : Managing Data Standards and Policies

You will find that as you begin to discover data quality rules you will need to adopt standards for things such as naming and coding conventions, permitted formats, valid events and other company data standards.

Historically these standards and policies may have been written in a document and shared in a folder that few people visited let alone implemented. Fortunately you can now link your standards to enforceable data quality rules in some data quality tools. This means that stakeholders can see for the first time whether these standards are being adhered to on a regular basis or whether certain teams or individuals are flouting the approved standards. At an operational level you can even provide real-time validation of standards and policies as data is being created or manipulated.

In one telecoms company for example we discovered several hundred coding conventions for equipment locations. This played havoc with procuring new equipment, provisioning spares and fulfilling support requests. By agreeing a company standard which was enforceable via the data quality tool and frequently monitored hundreds of times a day the quality of equipment coding rose dramatically.

Fortunately you can now link your standards to enforceable data quality rules in some data quality tools.

## Phase 5 : Data Quality Monitoring and Defect Detection

The earlier phases helped us ensure we have a clear view of:

- The various information chains that are used to support business functions
- The data quality rules that are in place across your data landscape
- The standards and policies required for good governance of your data

We have demonstrated how data quality technology can be highly beneficial for delivering these earlier activities but where data quality tools really outperform manual scripts or homegrown efforts is in the ability to continuously monitor vast amounts of data in search of defects.

In the past most organisations would detect defects on an ad hoc basis or perhaps during infrequent audits. Large amounts of defects can slip through the net with these approaches leaving the organisation vulnerable to service delays, dissatisfied customers and increased operating costs.

By using data quality technology we can monitor data quality levels throughout the data. We can even detect defects in real-time if we embed our data quality validation rules into the core applications and in-flight validation routines.

What's more, with this regular monitoring we can start to track data quality levels over time and across specific areas of interest. This allows us to see how different teams, data stores, processes, product lines and customer services are being impacted by defective data.

It is not enough to perform annual or bi-annual audits of your data. If you follow Phases 1-5 you will have a cost-effective data quality monitoring solution that runs regularly and delivers tangible returns 365 days a year.

By using data quality technology we can monitor data quality levels throughout the data. We can even detect defects in real-time if we embed our data quality validation rules into the core applications and in-flight validation routines.

## Phase 6 : Root-Cause Analysis and Defect Elimination

By now we have developed a robust system for detecting defects but of course we need to go further. We need to understand why defects are created and develop solutions for eliminating their cause.

We know from Phase 2 that Information Chains are a critical activity to manage in our goal for improved data quality. Information Chains support the core business activities of every organisation as data passes from function to function, often through many different systems.

In the recent Experian Data Quality research study, many respondents complained about the effectiveness of their root-cause analysis techniques. A lot of these problems will stem from lacking a complete view of the information chains in their organisation.

For example, if you manage a data warehouse and observe regular data defects then the only way to truly eliminate these issues is to map the information chains feeding the warehouse and work with upstream data providers to create long-term solutions. By tracking data quality along the information chain you can instantly observe where data breaks one of the constraints specified by your data quality rules. This is the benefit of adopting modern data quality technology, it allows you to create automated processes that can easily manage this level of maturity and complexity.

By observing these 'rivers of data' flowing around your organisation you can start to identify where people, process or technology changes are required to implement permanent improvements. Quite often it simply needs more education within the workplace so that data workers understand the impact of the mistakes they're making. Schemes such as bonus-related rewards based on data quality levels can also have a significant benefit.

Quite often you need to introduce an immediate solution to a data defect and a 'quick-fix' is required. You may be implementing a one-off data migration so you don't have the luxury of time to improve processes and re-educate the staff, you just need to cleanse the data in preparation for the migration. In this situation data quality tools are again a critical resource because they allow you to:

- Create automated cleansing routines for greater efficiency and scalability
- Provide a shared, auditable environment for non-technical staff
- Create full data lineage so you can see what other systems will be impacted by changes
- Monitor the impact of any changes made to ensure new defects are not introduced

Cleansing data manually is often costly, time-consuming and difficult to test adequately so once again data quality tools, this time with data transformation or cleansing abilities, come into their own.

In the recent Experian Data Quality research study, many respondents complained about the effectiveness of their root-cause analysis techniques. A lot of these problems will stem from lacking a complete view of the information chains in their organisation.

## Phase 7 : Data Stewardship

When speaking to organisations that have developed a mature data quality strategy it is clear that a system of Data Stewardship has been successfully adopted. Whilst the IT team can have a vital role in data quality management it is ultimately the business who is the true data owner. All too often the responsibility for fixing and managing data falls with IT and this is a mistake.

The business needs to take ownership and accountability for data and Phases 1-6 so far outline some of the core capabilities of a data quality strategy that can be implemented largely via business staff. IT have a supporting role but the onus is on the business to define data quality rules, internal processes, information chains, data standards and policies.

Data Stewardship is the amalgamation of all these earlier activities into an ongoing ownership framework. A business expert typically takes on the stewardship role and acts as a roving problem-solver working with the business, process, technology and customer facing units to resolve any defects found.

Data quality technology is of immense value to the data steward because it enables them to significantly scale up their area of influence. With appropriate tools they can manage huge amounts of data across multiple systems and services.

Data Stewards are able to monitor the data landscape and provide detailed reports for all levels of the organisation. For senior management Stewards can provide high-level trends and return-on-investment reports. For individual workers they can create personalised data quality stats. And for IT teams they can provide detailed recommendations for improving systems that are repeat data defect offenders.

## Section 5: Conclusion

The recent Experian Data Quality research study clearly shows that more organisations than ever are recognising the need for data quality management and improvement. However, many organisations are struggling to convert this desire for change into positive long-term results.

Some organisations have already invested in data quality technology and it is clear that in many cases the benefits have not materialised. Through countless interviews with data quality leaders it is clear that data quality technology can benefit organisations immensely but only if the technology is executed in tandem with the necessary data quality management techniques listed in this report. Fortunately, none of these data quality techniques are particularly complex or expensive to implement and the organisations that execute them correctly will observe major financial and operational benefits as a result.

The message is therefore clear. Long-term, financially viable data quality improvement is possible for organisations of any size and sector but only if they adopt a more robust data quality strategy that combines the necessary people, process and technologies.

## Section 6: Resources

- 'Global Data Quality Research 2014,' an independent market research report commissioned by Experian Data Quality and produced by Dynamic markets

## About the Author

Dylan Jones is the editor and founder of Data Quality Pro and Data Migration Pro, the leading community resources for their respective professions. He has over 20 years' experience of delivering complex data quality and data-driven initiatives. Dylan is a keynote speaker, author and regular publisher of expert insights on data quality related topics.

**Author**
Dylan Jones

Data Quality Expert
Data Quality pro

## 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 +61 3 8699 0100 | E info@au.experian.com | W www.experian.com.au/data-management