

# How to create a business focussed data quality assessment

Experian Pandora



**Created by Dylan Jones Editor Data Migration Pro.com** 

# Contents

Part 1: The Case for a Data Quality Assessment	03
Part 2: How to Implement a Data Quality Assessment	05
Step 1: Understand Your Profitability Drivers	07
Step 2: Build a financial performance model	10
Step 3: Create an Holistic Data Quality System	11
Step 4: Create an Interrogative Reporting Layer	12
Summary	14



## **Copyright Notice**

© Experian

This document contains information, proprietary to Experian, which is protected by international copyright law. The information contained herein may not be disclosed to third parties, copied or duplicated, in whole or in part, without the prior written consent of Experian.



## Part 1: The Case for a Data Quality Assessment

# 1.1 Why Do We Need Data Quality Assessment?

Organisations are finally understanding the potential value of high quality information assets. As a result, the need for data quality assessment techniques and technologies has risen sharply.

Whether you're implementing a new system, or simply looking to measure the quality of your existing data, data quality assessment has become a core activity that every organisation should master.

Data quality assessment is not an isolated activity of course. It should form part of a broader data quality management strategy but in many cases a thorough data quality assessment is one of the first steps organisations take on the road to improving and managing their data quality more effectively.

The goal of a data quality assessment is to answer the question:

"How closely does our data match our expectations, requirements and goals?"

A data quality assessment helps us to discover, document and analyse the gaps between where the quality of our data is and where it needs to be.

For example, we could ask questions of our data quality assessment such as:

- How is our data quality impacting service lead times?
- How is our data quality making us noncompliant with internal or external compliance directives such as Solvency II or Basel III?
- How is our data quality showing us that we're in breach of data protection or other legal requirements?
- How is our data quality costing us money or lost opportunities?

All of these deployments of a data quality assessment make the business sit up and take notice because if you can demonstrate to a senior executive that a strategic, tactical or operational goal will be impacted negatively by the quality of your data then they are far more likely to engage and support the ongoing data quality improvements you have discovered via the data quality assessment.

# 1.2 Where is a Data Quality Assessment Useful?

Here is just a sample of some projects and activities where data quality assessments are increasingly benefiting the organisation:



## **Data Warehouse Implementation**

In order for a data warehouse to serve a purpose, such as analysing historical sales performance, the data must be assimilated from across the organisation, transformed into the right structures and hosted over time for the business to gain value from the insights it contains.

There are opportunities for defective data to emerge at every step in the information lifecycle. Data quality assessments are therefore critical to understand the current quality levels of data and whether decision-making is being impacted.



#### **System Integration**

On a large scale data integration project such as a billing integration we want to know if our processes are working correctly so that we're compliant with regulations and also protecting our revenue.

In one project we discovered over £50,000 worth of unbilled revenue because of integration problems between billing systems and contract management systems. Where there are integration failures there are invariably performance or profit impacts and a data quality assessment is the perfect technique for discovering their impact.



#### **Data Migration**

Consider that most data migration projects fail through a lack of care and attention towards data quality management and you'll soon realise that data quality assessment should be a mandatory component of all data migration projects.

Bloor Research discovered that the the most common cause of failure on a data migration project is scoping issues followed by unrealistic timelines and poor data quality. Data quality assessment can help resolve all of those issues so little wonder why it is such a critical migration activity.



#### **Acquisition and Merger**

On a merger, we will want to verify that the company is being accurate with its statement of accounts or the reported number of customers it currently holds in its billing and operational systems.

For example, imagine your organisation is planning to acquire another company that claims to have 50,000 customers and a reported ARPU (Average Revenue per User) of £300.

What if the actual customer count is less? If there were a high volume of duplicate customers or duplicate households the ARPU figure would decrease and this could impact the value of the merger.



## Compliance, Risk and Fraud

For compliance we may want to check that a company has handed over all information of customers that meet certain criteria e.g. Terrorists/Criminals as part of a watchlist management process. Data quality assessment activities would form the backbone of this initiative.

In fraud investigations you would use data quality assessment techniques to assess whether there were irregularities in the way data is stored or transacted. Specialist data quality assessment tools can find hidden patterns and trends in the data that are unlikely to be discovered manually.



## **Master Data Management**

Organisations need to understand what data they have and where it is. They need to build an inventory of data assets with information such as what values exist, which of them are valid, are they consistently formatted, what are the unusual values, are there any implicit relationships between these values and others, is there duplication.

It is also important to understand where the same type of data exists in different systems, needing consolidation, as well as which data is duplicated across different systems and which is the best of the different versions.

All this information is required to allow feasibility, scoping, estimation and planning decisions to be made.



# Part 2: How to Implement a Data Quality Assessment

# 2.1 What Happens During a Data Quality Assessment?

We typically need some form of technology to assess data and create a series of metrics or criteria, for example:



17,121 customer records have a missing postal code



35% of install date fields in the equipment table are invalid



99.2% of financial transactions have a valid timestamp



300 out of 192,213 customer accounts are duplicates

So you can see how some metrics can be at the table column level, some metrics can be at the business fact level, some metrics can be at the record level - you can start to see how create different metrics for different audiences, functions and outcomes.

These metrics are extremely important because they give us a framework and direction over what to assess.

# 2.1.1 What are some common metrics (or data quality dimensions)?

Another term for metric is dimension and here are some of the most common data quality dimensions you will see in data quality publications.

So for example:

- Oniqueness is a measure of how many duplicate entries exist in an attribute
- Ompleteness is a measure of how many values (or records) are empty
- Onsistency is a measure of how many values are equivalent between separate data stores
- Accuracy is a measure of correctness against a reliable source or reality
- Validity is a measure of conformance to agreed syntax, formatting rules or other metadata definitions

Timeliness is a measure of how well data is made available during a specified time period

Agreeing on your metrics for assessment is very important and can be challenging because different authors have different approaches and terminologies.

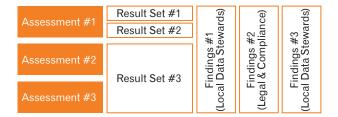
In my experience it really pays to invest in several books and study the articles on Data Quality Pro to get a broader understanding of dimensions and their usage because DQA's can be complex.

# 2.1.2 Data Quality Assessments Can Be Complex

When we create data quality assessments they can often serve multiple roles and functions of the organisation.

For example, we may perform multiple assessments across different areas of our data landscape so we may create different result sets depending on how the information will be used:

- Local data stewards may want to know which individuals are breaching local data policies and standard operating procedures.
- Legal and Compliance staff may want to know how data quality is impacting corporate compliance rules or external directives such as Basel III and Solvency II
- Business leaders may want to understand how poor data quality impacts their strategic goals such as competitive advantage or customer service improvement



# 2.1.3 The Great Data Quality Assessment Challenge

One of the problems with data quality assessment is historically it was seen as a data auditing task requiring technical skills and therefore firmly the responsible of the IT domain.

Due to the techno-centric nature of these assessments the results often lacked relevance to the business community who would struggle to

take any action because the results lacked focus and prioritisation of business issues.

Where assessments focus too much on basic technical dimensions such as uniqueness and completeness then the business often respond with a 'So What?' attitude. Because even basic data quality tools can cope with measuring simple dimensions like these then it was invariably the first effort taken by organisations on their journey to better data.

Without clear business relevance the business fails to get involved and as a result the data quality initiatives can fall at the first hurdle.

# 2.1.4 But I thought our data quality tool was meant to help?

Having an effective data quality tool that supports data profiling and data quality assessment is still an incredibly useful asset to have, you just have to think beyond the basic technical checks that so many practitioners fall back on.

You need to leverage your data quality functionality to tell a bigger and better story by integrating business metrics. Data quality tools should provide all the benefits of automation, scalability and a business-friendly interface thus allowing data quality assessment to become very much a business-driven activity.

# 2.2 How to Improve the Traditional Data Quality Assessment Approach

We've seen how historically many companies have taken a techno-centric view of data quality and data quality assessment in particular so how can we improve the situation?

The approach I typically use goes something like this:



Step #1: Understand Your Profitability Drivers



Step #3: Create an Holistic Data Quality System



Step #2: Build a financial performance model



Step #4: Create an Interrogative Reporting Layer

# 1 Step #1: Understand Your Profitability Drivers

We need to understand the profitability drivers of the organisation. Where does it create revenue, where does it incur costs. How is that changing over time? Which products and services create the most profits? What is the lead time for each service? What are the process steps of each service and where does the cost occur at each stage?

# 2 Step #2: Build a financial performance model

This is where we pull in a lot of financial and performance related information to create a much clearer view of how the organisation is performing in our area under assessment. This could be a department view, product view, service line - wherever we've been tasked to perform a data quality assessment.

# Step #3: Create an Holistic Data Quality System

We need to build an Holistic Data Quality System that incorporates not just data quality metrics but financial and performance too from Phases 1 and 2. Solely focusing on technical metrics and basic data quality measures can alienate the business so we need to connect data quality metrics that allow the business to drive real change and improvement.

# Step #4: Create an Interrogative Reporting Layer

With an interrogative reporting layer we can let the business loose and give them the ability to ask business questions of the assessment results instead of spoon feeding them data profiling stats that can often serve them little purpose.

It is essential to give the business access to both high-level summary and low-level detail to allow them to focus their efforts as well as understand and resolve issues.

# Step 1: Understand Your Profitability Drivers

To explain this step let's start with a fictitious example taken from the telecoms sector.

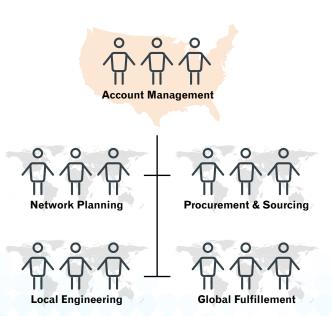
All telecom providers sell businesses communications solutions so they need to ensure they fully optimise their internal cost centres, recover all client revenues and ensure service level continuity.

In our example we have a US investment bank that wants to create a new global communications infrastructure to upgrade their connectivity with regional offices, as we can see in Sydney, Australia in the next image.

The business leader asks their telecoms provider to make this happen.



What follows is a series of processes initiated internally by the telecoms provider:



The telecoms firm initiates many processes and people need to be deployed for planning the network, procuring new fibre links, shipping equipment out to the local sites and tasking an internal or external engineering team to connect it all up and get it operational.

In reality there are a lot more processes than this but it shows you just how complex telecoms services can be. It also illustrates just how much information is required for a process like this.

The following image on the next page is an example of the engineering process involved just in switching on a new customer. It doesn't even include a lot of the teams we saw on the previous page.

As you can see it's quite involved and depends on multiple systems and obviously lots of different data sources so lots of opportunity for poor quality data to impact the profits and performance of the business.

We could periodically assess the quality of data across each of these different systems and teams but we want to do more than this, we want to create a business focused data quality assessment so we go one step further. We start to understand the business model of this process and where profits are created. To do this we need to break down the costs and revenues involved.

We need to look at things like cost of equipment and internal services as well as all the revenues generated from the service. We can do this at a high level because department heads and functional heads will know their quarterly burn rate and revenue rate but they'll often lack a lot of granular financial metrics that are locked away in the data.

Here are some of the financial metrics we would like to know:

- Ost of installing 3rd party equipment
- Ost of leasing international lines
- Ost of hiring contractors or local workers
- Ost of lead generation and sales
- Revenues generated from service

In addition to the financial metrics it will also help our data quality efforts if we understand performance related metrics such as:

- Ocst of installing 3rd party equipment
- Ost of leasing international lines
- Ost of hiring contractors or local workers
- Ost of lead generation and sales
- Revenues generated from service

# What If You Don't Have Access To Business Metrics?

Chances are that these type of metrics may not be readily available in your organisation, at least in a format you can use.

You may face objections such as:

- Some of the metric data belongs to a different owner
- Some of the metric data is in different systems
- Some of the metric data is in paper records
- Some of the metric data can't connect
- Some of the metric data is missing

These objections are to be expected and many practitioners will walk away at this point and simply return to the traditional, techno-centric data quality assessment.

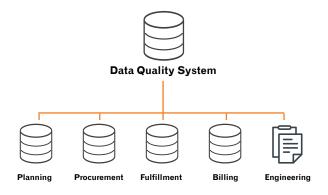
However, you should persevere because objections are good.

Lack of quality metrics that the business values demonstrates a case of poor information quality. This gives you a perfect opportunity to add value and demonstrate that your data quality team are not just there to create a glossy report but to actually help the business become more efficient and effective.

You can use the same data quality functionality that you need to for the data quality assessment in order to start pulling disparate pieces of information together.

This enables you to understand the profitability drivers of the process, department, information chain or any other scope you have for the assessment.

The next image demonstrates how your data quality system can be the hub that connects disparate systems together in order to gather metrics that were previously unobtainable.



You can start to create a consolidated view of these different systems and when you do this you naturally build up a more detailed account of how the company creates revenues and incurs losses i.e. your profitability drivers.

I need to emphasize this point again. When you can link data then you understand profitability far more clearly because how a company thinks they're performing is very different to reality when you start to get more granularity in your metrics. For example, a network company was basing its financial performance on the electronic contracts system it used which had literally thousands of very small payments listed to 3rd party contractors across the globe.

When we compared this data to the actual billing statements we found major discrepancies. What was on the electronic contracts system was often garbage because the telco was still issuing monthly cheques to contractors who had terminated their leases months or even years before. In a huge corporation, these can add up to significant sums.

But of course it's not just big companies who are struggling. I published a story last year of a small company (literally a few employees) who narrowly avoided bankruptcy because one of their staff was submitting fraudulent invoices. What we have discussed in terms of using data quality tools to link disparate systems is exactly how fraud

investigators work. They 'stitch' together different sources of data to find anomalies that lead them to financial irregularities.

If you want to carry out business focused data quality assessments you need to think about linking as much revenue, cost and also performance information into your data quality environment as possible.

In the following diagram we can see various linkage points with other systems that can help us pull in additional data to add relevance and context to our operational data. This is how you start making data quality assessments more engaging and purposeful for the business.



Links to other systems, you can enrich with financial and performance data Performance Metrics and Data Quality Metrics

# Step 2: Build a Financial Performance Model

In Step 1 we started to think about the various costs and revenues associated with our scope of assessment and where those information sources lie.

We now need to pull all this information together and create a financial performance model so that we can start the process of linking data quality metrics.

As we walk through the business process under review we start to identify these pockets of data that are useful to our assessment. Where in the past, using a basic data quality assessment, we may just examine these systems in isolation, we now need to look at linking them into our overall financial performance model.

In Step 1 we're really trying to understand what the profit drivers are that we want to know. Step 2 is focused on pulling this information into a staging area where we can manipulate it further and make the information more useful. We're still carrying out a physical data quality assessment of the connecting information of course because we want to be sure that we're not introducing any errors.

A good example of this involved some work I carried out for a telco that was looking to improve field force efficiency. They had a hunch that data quality was to blame but after a fairly extensive data profiling exercise they couldn't find any major issues.

I initiated Step 1 and started to discover all of the financial and performance related information available. I created a staging area of staff salaries, another database of field service logs and another of equipment planning and provisioning.

In the field service log it showed the start time and end time of a task and which engineers worked on a task. I was then able to calculate the duration and cost of each job. No-one had done this before. We then looked at the equipment planning system and could see where revisions were being made and a lot of these were done because of errors during the planning stage or because field force workers were not recording field data correctly.

Just by 'mashing up' different financial data with data quality metrics in this way we were able to get a much clearer view of financial impact. This is precisely the type of data quality impact assessment business people need to take action.

### But Isn't This Just Business Intelligence?

You may well be reading this and thinking to yourself: "Mashing up data, integrating financial and performance metrics - isn't this just business intelligence?"

You're absolutely right.

Ironically, it's a direction where some modern data quality technology tools increasingly seem to be heading. They can actually integrate data quality and financial metrics into the tool itself. With all the dashboards and interrogative functionality already built in some data quality tools are starting to resemble business intelligence environments more and more. Some even use the words 'Data Quality Analytics' or 'Data Intelligence'.

We measure data quality metrics and business performance metrics in essentially the same way. Calculations are made at a granular level and aggregated up and across an area of the business that matters to us.

What I'm outlining in this report are some new ideas around how to bring them into the same environment so we can engage and focus the business.

## Step 3: Create an Holistic Data Quality System

**Holistic:** adjective: "characterized by the belief that the parts of something are intimately interconnected and explicable only by reference to the whole".

Why bother with an holistic data quality system?

Historically, a lot of data quality assessments are very technically oriented, taking a system by system, table by table, column by column approach. Essentially this is measuring data quality in isolation.

A lot of data owners will state that their data in these cases is absolutely fine, with no reported issues.

In many cases this will be true, their data may be fine for their local processes or when assessed in isolation but the problem is that if you just measure data locally then it won't give you the full story.

**Data isn't an isolated asset.** Data must form connected information chains that span across multiple departments, systems and even organisations, all as part of the vital organisational processes that drive revenues, incur costs and deliver services.

This connectedness of data in support of complex, multi-application business functions, means that a more holistic data quality management approach is required.

Let's take a look at some classic examples to illustrate the importance of holistic data quality.

#### Assessing Completeness

Completeness in the [Planning.Inventory] column tells us if there are any missing values but...

Value	Complete?
Juniper M320	YES
Juniper M320	YES
Juniper M312	YES
Juniper m320	YES
Juniper (M320)	YES

... it doesn't tell us if there are other values that should probably be there from

Master Stock Item Juniper M320 Juniper M312 Juniper M315 Juniper M325 Juniper M327	Quantity 10 — 07 — 03 02 15	<b>'</b>
Value  Juniper M320  Juniper M320  Juniper M312  Juniper m320  Juniper (M320)		Complete? YES YES YES YES YES YES YES

## Step 4: Create an Interrogative Reporting Layer

By this point you will have a consolidated view of your:

# Financial Metrics + Data Quality Metrics + Performance Metrics

However, we need an environment to merge all this information together in order to ask the right questions about our data and present findings in a context that the business (and technical) community will understand.

# Introducing an interrogative reporting architecture

The following image provides an overview of the architectural elements of an interrogative reporting architecture:

Dashboards and Visualisation				
Data Quality Assessment Rules	Data Profiling and Data Discovery	Data Repository and Storage		
Information Chain Management	Data Quality Monitoring and Alerts	Data Integration/ Movement		



#### **Dashboards and Data Visualisation**

Modern data quality tools feature a great deal of visualisation capability and of course with open repositories comes the opportunity to integrate your data quality performance metrics into existing business intelligence solutions that most organisations now possess.



#### **Data Quality Assessment Rules**

These are the core of your data quality assessment strategy. These rules define what 'quality looks like' from your data. These can be as simple as basic rules to check for NULL or empty values through to very complex rules that check for a variety of conditions and constraints with data coming from multiple systems.



## **Data Profiling and Data Discovery**

These functions are critical for automating the discovery of data quality assessment rules and without these it's virtually impossible to scale up a data quality assessment effort. Beyond this initial discovery phase these functions are also useful for continually discovering new rules, validating findings and ensuring that existing rules are still valid.



#### **Data Repository and Storage**

Your data quality assessment should have as little impact on the operational environment as possible. You also need to have access to the detailed data behind all your measurements. It is therefore advisable to ensure you have a repository or storage strategy in place for your data during the assessment. Modern data quality tools have a repository built in to the assessment environment but it also helps if you are able to track the performance of specific rules and measurements over time. To do this you will need not just basic data storage but a data quality repository so that you can spot trends in the rule measurements over time.



## **Information Chain Management**

As we read earlier, data quality assessments of individual columns and tables can be useful but will fail to measure data quality levels across your organisation. To measure holistically we need to discover and manage information chains. This requires automated relationship discovery functionality and also the ability to create complex data quality rules that validate and measure quality across many systems simultaneously.



## **Data Quality Monitoring and Alerts**

Once your data quality assessment framework is in place it is extremely helpful to have your data quality rules continually measuring quality levels and reporting on any violations found. This type of monitoring is more operational data quality and relevant for local data stewards and data workers who need to take immediate action or make process improvements.



## **Data Integration/Movement**

There needs to be some mechanism for ensuring that the repository is regularly refreshed with operational data so that assessments can be repeated and trends analysed. Data movement is therefore critical but integration with a wide variety of data sources and database formats is also highly beneficial.

# What are the benefits of an interrogative reporting environment?

A lot of data quality assessments fail to make an impact with the business because they talk in terms of data statistics and quality metrics, terms that the business simply don't understand. We have shown how it is possible to build an environment where business metrics, performance metrics and data quality metrics can all be supported, connected and visualised.

This connectedness and visualisation process allows us to answer questions such as:

- How long does it take to complete an order?
- What staff costs are involved in fulfilment?
- What costs are incurred when data quality errors force a re-design?
- What is the frequency of delayed orders over the last 12 months?

These are the sorts of questions that the business wants answers to and that we can now provide with our consolidated reporting layer.

By adding performance data and financial data to our data quality measures we can start to create a great deal of focus around the most profitable areas of the business to improve data quality.

For example, one telecoms company found their fulfilment processes were taking longer and were increasing their stock of unplanned equipment.

The issue actually lay with a rogue software release that introduced defects into the information chain.



## Stage 1: Errors Introduced

Software developer introduces a new piece of functionality that searches for product codes. It fails to find existing codes in a certain product line.



# Stage 2: Errors enter the information chain

Users begin ordering new equipment and fail to return existing equipment back to spares and re-allocation, error begins to cause further downstream defects.



## Stage 3: Holistic reporting environment

Using a consolidated reporting architecture the telco is able to observe rising costs, longer lead times and data quality defects. All the information is to hand in order to analyse the cause and effect of the issue.



## Stage 4: Alarm raised

By setting thresholds on a variety of data quality and other metrics the relevant stakeholders can be notified of any issues. In our case a software change was implemented to prevent a repeat of the problem.



## Summary



**Takeaway #1:** Stop thinking of data quality assessment in terms of isolated analysis. For your data quality assessment to truly have an impact on the business you need to stop assessing individual tables and even systems. Instead build a complete 'assessment ecosystem' combining financial, performance and data quality metrics.



**Takeaway #2:** Understand how your business model operates and create an end-to-end view to model it, identifying data sources along the way. There is far more information you can store in this model but unless you fully understand how your organisation uses information to create services and products then you'll never be able to transform data quality assessment results into solid business change.



Takeaway #3: Focus on the goal of your data quality assessment and create an architecture to support that objective. Every organisation has a pressing agenda so find out whether your sponsors are concerned about lead times, customer satisfaction, profits, costs, market share or other drivers. Make sure you are able to build an assessment architecture that models and measures the systems, functions and data that align to drivers the sponsors care about.



**Takeaway #4:** Don't just spoon-feed the business with canned reports. Instead, give them an environment that they can interrogate. Let them discover the issues that matter to them most. Data browsing and visualisation are key here.

## **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.

Experian Australia Pty Ltd Level 6, 549 St Kilda Road Melbourne, VIC 3004 T: (03) 8699 0100 info@experian.com www.edq.com/au



