experian™

# How to take control of data quality on your next data migration

**By Dylan Jones | Editor of Data Quality Pro & Data Migration Pro**

# Contents

**About the Author**

Dylan Jones is the founder of Data Quality Pro and Data Migration Pro, popular online communities that provide a range of practical resources and support to their respective professions.

Dylan has an extensive data migration and data quality background and is a prolific publisher of expert articles and tutorials on all manner of data related initiatives.

# How to take control of data quality on your next data migration

## Introduction

It was all going so well.

The test migration of sample data went through with only a handful of minor issues. The target application functioned as expected. The business signed off their use case tests.

There was a green light for the go-live weekend, and all was going to plan until late on the Sunday evening when the problems started. Poor data quality began to rear its ugly head.

First, the load processes kept churning out failed records.

Incorrect formats and broken relationships, amongst a host of other defects, caused over 5% of records to fail.

"Why didn't we spot these issues in our test data, we thought we had representative samples?" people begin to mutter amongst themselves.

To compound problems of missing data in the target, the records that did load suddenly started wreaking havoc with the new billing system. The business users test their monthly billing cycle, and it fails in spectacular fashion.

People begin to ask: "Where did it all go wrong?".

For the uninitiated, the previous anecdote is a typical encounter with data migration projects when data quality management is omitted from the overall strategy.

But there are promising signs within the data migration sector that best-practice methods and technologies are increasingly being adopted.

Recent research by Data Migration Pro (commissioned by Experian) discovered that 69% of recent data migration projects were considered a success, with 72% of projects being delivered with less than a 3-month overrun.

One of the reasons for a successful outcome was the fact that more organisations than ever are deploying data quality tools and methodologies during the migration process.

For example, where an 'excellent' data quality strategy was implemented, 67% of projects completed without any overrun. In contrast, when the data quality strategy was classed as non-existent, poor or very poor, only 37% of projects completed without any overrun.

So what elements should you include in your data quality strategy for a data migration?

**This guide provides some practical steps to ensuring a smooth delivery of data quality services on your next data migration.**

---

**Steps discussed are:**

**Step 1:** Understanding the Path Ahead

**Step 2:** Forming a Data Quality Alliance

**Step 3:** Defining the Data Quality Management Workflow

**Step 4:** Implementing the Data Quality Architecture

---

## 69% of recent data migration projects were considered a success, with 72% of projects being delivered with less than a 3-month overrun.

**– Data Migration Pro (commissioned by Experian)**

# How to take control of data quality on your next data migration

## Step 1: Understanding the Path Ahead

You may think that your data migration is straightforward.

Your IT suppliers and partners may convince you that for them this is just another migration, one of many success stories.

But why take a chance? Why let guesswork and assumptions negatively impact your project (and career)?

Too many project leaders have fallen into the trap of hoping for the best with data quality with a 'let's fix it in the target' mentality.

It doesn't need to be this way.

For the project manager, Data Quality Management should be an ally, not some combative foe they need to avoid.

Applying sound Data Quality Management on a data migration can help direct the strategy and even the architecture of the project because it will discover hugely important facts about your data that have long been forgotten.

For example, a large European utility organisation wished to migrate their entire equipment database to a new, 'state of the art' target application. They purchased the target system and commenced what was assumed to be a trivial task of migrating the legacy data across.

They instantly hit roadblocks.

First, they discovered that some of the critical systems contained equipment that had security restrictions so progress was suspended until it could be safely encrypted.

Then they found the target system required a far more stringent set of quality constraints than the legacy system. This 'data quality gap' meant that the legacy data that was missing or inaccurate would have to be rectified before the migration.

Finally, they realised that, due to the way the data was structured, the quality of the data meant that in its current state it was impossible to reliably link three of the core systems that were central to the target implementation. The reality is, had they applied data quality management upfront, all of these problems could have been foreseen, planned for, and overcome.

## Applying Data Profiling and Discovery (Look Before You Leap)

Never embark on a data migration project without fully discovering and assessing the quality of your legacy data. This process will help you learn so much more about how your data has been utilised in the past and the pressures it will face in the future target system.

For this reason, it is recommended that all migrations be subjected to a 'Pre-Migration Impact Assessment' (PMIA).

A PMIA is a rapid investigation and analysis activity that gathers vital information such as:

- Wide scale data discovery of legacy systems to uncover relationships between disparate data sources and identify probable missing/duplicated data

- Extensive data profiling to create a definitive catalogue of metadata and initial data quality rules base

- Discovery and prioritisation of data quality issues so that a pragmatic strategy to the data migration can be undertaken

The PMIA is an activity that focuses on the use of data profiling, data discovery and data quality tools to replace assumptions with knowledge prior to the migration commencing (as seen in the following diagram).

**Diagram 1: Removal of assumptions during a PMIA**
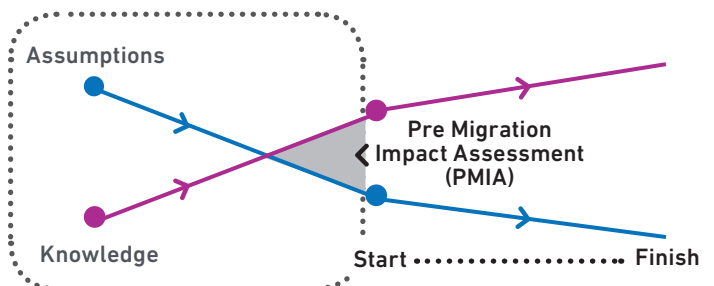


Despite the obvious benefits of discovering the risks and pitfalls of a data migration before commencing, a recent research initiative commissioned by Experian found that only 36% of organisations delivered a thorough impact assessment on their last data migration.

You can find out more about Pre-Migration Impact Assessment in several resources at our Data Migration leadership hub:

https://www.edq.com/uk/data-migrations

"The client simply didn't understand their source architecture at all, they completely underestimated the scale of the project and I think it's fair to say the target system supplier didn't help, they made it all sound so simple. The problem is [the organisation] didn't have a lot of experience with this type of migration because they were a non-profit with limited technical expertise in the department. Would an impact assessment have helped them? Undoubtedly, because there would have been no more assumptions and guesswork from the target vendor."

— **Comments by a research respondent, Data Migration Pro Research Study 2017**

## Step 2: Forming a Data Quality Alliance

Once the project starts, your next task will be to assemble the team that will manage data quality for the duration of the project. This is another stage that many companies get wrong by ignoring the need for a cross-functional group, drawing on technical, business and supplier-side personnel.

It's not uncommon for data quality management to be staffed entirely by the systems integrator or contractor responsible for the data migration. The danger with this approach is that it's the business who owns the data, they are its true custodians, not a Database Administrator or backoffice IT team who may not understand the true meaning and context of the data.

If you want to manage data quality effectively you need to get involvement from across the organisation.

**Who should you invite to your data quality alliance?**

We cover the topic of building a data quality workflow in Step 3 but here are some examples of the data quality roles involved and where these resources can be sourced from:

**Data Quality Workflow Leader:** On a large project, you will have a designated leader of the data quality aspects of the data migration. This frees up the data migration experts to focus on other areas, safe in the knowledge that data quality is being coordinated. On smaller projects, the data quality team leader role will typically be bundled into the data migration lead. Ultimately it needs someone who can make decisions and has a strong understanding of data quality management methodologies and software.

**Business Subject Matter Experts:** These will be the 'Gold Users' of the business world. Those savvy souls who know all the idiosyncrasies of their legacy system and can help unravel the meaning of long-forgotten datasets.

**Technical Subject Matter Experts:** These will consist of legacy system experts and target system experts. Ideally you want representation from software developers and designers, database administrators, data architects and interface designers. These are the people who will help you understand the various structural architectures and design.

**Data Quality Analyst:** Large projects may benefit from dedicated team members who can execute data quality assessments, data profiling and data discovery tasks. These people are the 'engine-room' of the data quality process. They find issues, present them to the team for resolution and even design creative fixes to prepare the data for migration.

"Business engagement on data migration projects has been proven to be a key driver for success but we found that half of all projects surveyed lacked adequate business engagement on the data migration. Interestingly, we found that where projects had overrun significantly (beyond 1 year) only 21% of projects had full business engagement. This appears to show some form of correlation between engagement and project outcome."

— **Findings from the 2017 Data Migration Pro Research Study**

## Step 3: Defining the Data Quality Management Workflow

You now have your data quality 'dream-team' but how should they work together?

What systems and processes must they follow?

What you need to avoid is the situation where everyone on the project 'does their own thing' with data quality in an uncoordinated manner.

To get around this common problem you need to create a formal process that follows a simple flow like the one in the following diagram:

**Diagram 2: Data Quality Workflow for Data Management**
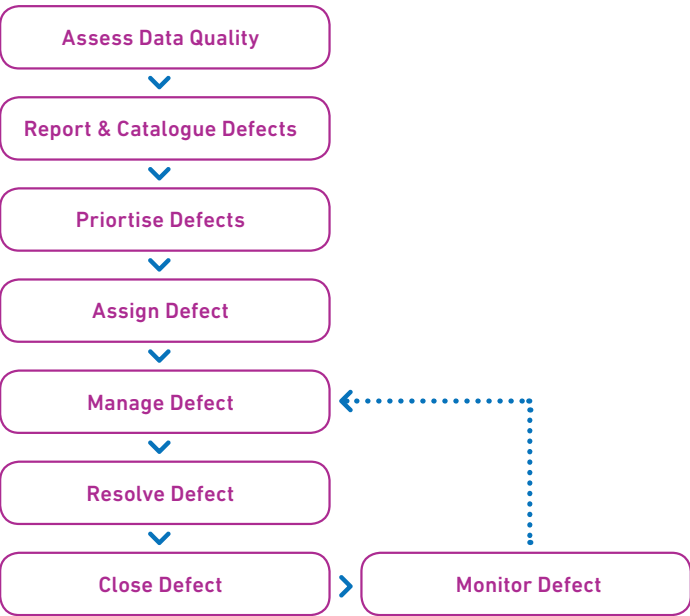


Despite its simple design, there are some key points to consider in this diagram:

**Assessing:** Defects are discovered via a data quality assessment phase and the normal discovery of issues picked up during design, build and testing. It's important to have a robust data quality assessment strategy so that you're not using other phases (particularly testing) to uncover defects.

**Monitoring:** The monitoring of defects continues even after the defect has been closed. Many data quality issues can be the result of bad information sources that can easily be repeated (such as poor data entry or upstream data feeds). You therefore need to keep the data quality monitoring processes running to ensure no repeat failures are creeping back in.

**Prioritising:** Prioritisation is critical. Make sure you are not just passing defects over to your resolution staff without assessing the importance of each defect. Some issues can be more easily resolved in the target system (where applicable) and conversely some issues are critical to the success of the project so need to take precedence over everything else. Some issues may be frustrating but not deemed critical, so the resolution as far as the migration project is concerned is 'do nothing'.

**How are staff structured around the data quality workflow?**

We introduced the different resource types that we need on a data quality team in the previous section but now we need to assemble them into a team.

First, we need a 'Data Quality Council' and this will typically be led by the data migration project leader. They may defer leadership once the workflow is set up and running but it is the project leader's responsibility to ensure it is initiated correctly.

The Council Administrator will work in tandem with the project leader to assign priority of the defects discovered. This person will ideally be a business representative as they should have the context of whether the defect really is critical to the business.

The rest of the council will consist of technical and business experts combined with the data quality analysts. They will meet once a week (or more regularly depending on project size) to review all the issues under management and report on progress.

In terms of managing, monitoring and closing defects, this activity is coordinated by the data quality analyst team. They may call in the support of the business to fix the defects at source, they may devise cleansing routines in the data quality software to fix the data in-flight during the live migration or they may create an operational note to fix the data in the target system.

Additional resolutions may be required depending on the tools and architecture available but it is the role of the data quality analysts to accurately record the scale and progress of each defect and ensure that it is being monitored right up to the point of migration.

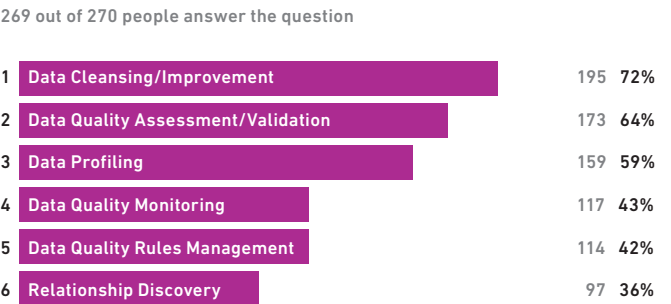## Step 4: Implementing the Data Quality Architecture

Given the importance of data quality, there is a clear need for the appropriate use of data quality management technology on data migration projects.

For example, you simply can't hope to deliver a robust data quality strategy with hand-coded analysis and manual cleansing given the huge volumes and short timelines witnessed on a modern data migration project.

This need for data quality technology was confirmed in the recent 2017 Data Migration Pro Research Study with 72% of projects using data quality cleansing or improvement software and 64% of projects using some form of data quality assessment or validation software.

The other major data quality software components observed are included in the following diagram:

**Diagram 3: Select all the data quality software functions that the team used on the project?**

269 out of 270 people answer the question

| | | | |
|---|---|---|---|
| 1 | Data Cleansing/Improvement | 195 | 72% |
| 2 | Data Quality Assessment/Validation | 173 | 64% |
| 3 | Data Profiling | 159 | 59% |
| 4 | Data Quality Monitoring | 117 | 43% |
| 5 | Data Quality Rules Management | 114 | 42% |
| 6 | Relationship Discovery | 97 | 36% |

The image below outlines some of the data quality capabilities that you will need to cater for, particularly on the larger scale data migration projects.

**Diagram 4: Typical Data Quality Management Architecture Required for Data Migration Projects**
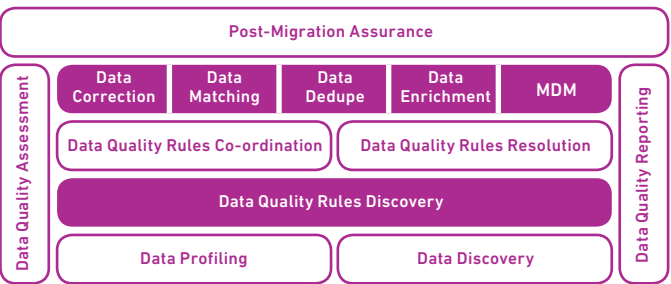


**Data Profiling:** This enables the data quality analyst team to automatically discover the data quality rules found within the legacy data. Over time the documentation and knowledge of the legacy data can be lost so data profiling enables an accurate and current knowledge-base of the legacy data to be formed. Data profiling also helps establish how the data is being used and whether there are any obvious defects present.

**Data Discovery:** Typically carried out in tandem with Data Profiling functionality. Data Discovery is the ability to automatically uncover all known relationships between data sources and assess their quality. On a data migration, this is a vital function because the legacy data landscape can often be disparate and poorly understood. Data Discovery helps you build all the mappings and relationship rules that you need to begin the design process. It also allows the scope and estimated complexity to be validated (or not!). With an automated data discovery tool, you can shave many weeks off your development timescales.

**Data Quality Rules Discovery:** Some of the more advanced data quality tools will actively discover and document the data quality rules that are present in your legacy data. This can be due to data profiling or data discovery tasks but also by techniques such as functional dependency analysis which automatically finds implicit data consistency rules that can take far longer to uncover using manual approaches. User workshops are also vital for discovering the functional makeup of the legacy and target environments, then translating these into data quality rules via your data quality tool.

**Data Quality Assessment:** Once you have determined the scope of your data migration and discovered your data quality rules, you can commence the data quality assessment phase. This provides an exhaustive, qualitative review of all the legacy data that will support the migration. A modern data quality tool can possess several hundred data quality benchmarking metrics that can validate everything from hidden patterns in the data through to expected ranges and business rule conformance. When combined with information about priorities, this determines where the Data Quality effort will be spent. It is the most critical resource in the entire data quality architecture stack.

**Data Quality Rules Coordination:** The entire data quality management workflow is driven by data quality rules. The data quality rules coordination process will track all the defects discovered and their appropriate 'action plan' for resolution. Your 'data quality toolkit' should enable you to coordinate the resolution of these rules by assigning them to different team members, reporting on their resolution, re-opening past issues, flagging their importance etc.

**Data Correction/Cleansing:** Where data is found to be defective, you will want to apply a range of resolution strategies to resolve the data either at source, during migration or even in the target environment. Your data quality tool should be able to 'cleanse' or fix the data using specialist transformations. Sometimes you may elect to do this prior to the migration so that the legacy data can be used by legacy business users. Other times you may decide to simply correct the data as it flows through the migration architecture during the run-time phase of the project. Obviously, you need a data quality tool that can cope with these different arrangements.

**Data Matching:** A common requirement on data migration projects is the ability to link disparate datasets together. If your data quality architecture possesses this feature you may not need to purchase a more expensive ETL tool for example, the data quality architecture can serve the entire project. Data matching is very useful for supporting the deduplication of legacy data, for example where equipment or customer records have been replicated through user or system error, or are coming from different systems.

**Data Deduplication 'Dedupe':** Working in tandem with the Data Matching functionality, Data Dedupe functions enable you to prepare the legacy data for the live environment without necessarily impacting the legacy operational systems. Quite often you are not able to clean up the legacy data for fear of breaking existing processes and existing feeds to other systems (e.g. data warehouses). By having a dedupe function on your data quality architecture you can easily resolve duplicate data during the live migration without impacting any of the existing business functions.

**Data Enrichment:** Quite often we need to pull in 3rd party data sources to enrich or extend legacy data to ensure that the target system has a complete dataset. For example, you may find that your legacy equipment records are lacking manufacturer information but contain the product type code. By matching the product type code to a 3rd party list you can enrich the data with a valid source of information so that you can now populate the manufacturer name, power ratings and so on. This function may work in tandem with the data matching functionality to enable the linkage to take place.

**MDM (Master Data Management):** On a data migration project, we won't be looking to implement full-blown Master Data Management but we need to ensure that our master entity data is being controlled appropriately. What we don't want is for developers to create their own sources of legacy master data (e.g. equipment types, customer types)

for example. Your data quality architecture can help here by creating a full audit trail, configuration management, deduplication and matching service to ensure that your master data is populated once but shared wherever necessary across the project.

**Data Quality Reporting:** You need functionality that reports on the health of data quality defects so that we have a 'readiness' score for the migration. Since the business can see the 'broader picture' beyond the confines of our migration project, they may decide to migrate even though there are unresolved issues. However, they can only do this based on the objective measurement of the data. Reporting is particularly useful for the Data Quality Council so that they can see the progress of team members tasked with resolving defects. Without a reporting function the whole team is 'flying blind'. Without regular monitoring and reporting, it is also difficult to spot when defects have reoccurred.

**Post-Migration Assurance:** One of the key benefits of creating a best-practice data migration project that adopts the right strategy for data quality management is once the migration completes you are left with a fully functional data quality process and architecture that can be handed over to the business and IT community for long-term assurance. All data will inevitably become defective if not well managed so it is critical that your new target environment be adequately governed. By investing in the right data quality tools and resources during the migration you can easily retain that resource and ensure the future success of your target system.

## Summary

Data Quality Management should form the backbone of your data migration strategy.

Look at your data migration strategy and project plan - have you included the steps covered in this guide? Where do you need to adapt your approach to cater for increased data quality management?

The typical arguments against data quality are that it costs too much and takes too much time but the irony is that effective data quality management actually reduces the duration of a project and improves its likelihood of success.

What's more, once you've successfully applied data quality management during the data migration, all the technology and methods can be retained for delivering ongoing data quality assurance and data governance benefits after the data migration. Many companies have adopted this approach as a catalyst to developing a more robust and mature approach to data quality management across the enterprise.

## For more information contact us:

Telephone: 0800 197 7920
Email: dataquality@experian.com

## Next Steps

### Want to take control of data quality on your next data migration?

Experian Pandora delivers a complete data quality and data migration solution.

With one platform, you can now deliver an end-to-end capability that encompasses all the techniques presented in this guide.

In addition, you get hundreds of additional data migration and data quality functions to help empower the entire team, from Pre-Migration Impact Assessment through to Legacy Decommissioning and Post-Migration Assurance.

Book a free consultation with one of our skilled data migration and data quality practitioners to talk through your project roadmap and learn how Experian Pandora can accelerate and de-risk your upcoming data migration.

## Resources

To help you create a best-practice data quality management process on your data migration we have assembled a broad range of resources at our Data Migration Leadership Hub:

https://www.edq.com/uk/data-migrations/

## About Experian

Experian unlocks the power of data to create opportunities for consumers, businesses and society.

At life's big moments – from buying a home or car, to sending a child to college, to growing your business exponentially by connecting it with new customers – we empower consumers and our clients to manage their data with confidence so they can maximise every opportunity.

We gather, analyse and process data in ways others can't. We help individuals take financial control and access financial services, businesses make smarter decisions and thrive, lenders lend more responsibly, and organisations prevent identity fraud and crime.

For more than 125 years, we've helped consumers and clients prosper, and economies and communities flourish – and we're not done.  Our 17,000 people in 37 countries believe the possibilities for you, and our world, are growing. We're investing in new technologies, talented people and innovation so we can help create a better tomorrow.

**Learn more at www.experian.co.uk**

**Experian**

**T** 0800 197 7920
**E** dataquality@experian.com
**W** www.edq.com/uk