White Paper Series : Business Guide to Data Migration

White Paper #1 - Data Profiling & Remediation

About the Author

COSOL is an Australian based organisation with 20+ years of experience in complex ERP system upgrades, migrations, and consolidations. Through this experience COSOL has developed a strong reputation for end-to-end data migrations which remains at the very core of COSOLs expertise.

This paper is the first in a series summarising lessons learned, to provide insight for executives who are faced with a possible large-scale program of this nature within their own organisation.

Our Point of View

Experience shows that one critical success factor for these programs to realise the value in their business cases is Data Migration. The axiom "garbage in, garbage out" remains as relevant today as it did when computers were first invented. COSOL strongly believes:

- that **digital transformation is well underway**, and every Board is, and should be worried about how to become a truly digital enterprise.

- **strong Enterprise Data foundations will be required** to enable adoption of digital solutions including advanced analytics, robotic process automation, machine learning and artificial intelligence which are **the next frontiers to productivity** and market competitiveness.

- Enterprise Data is the glue, the fact base, that **drives decision making and business improvement**, allowing organisations to meet stakeholder expectations in a timely and efficient manner; and

- for organisations to succeed, **Data must be treated as a mission critical Asset**; It is the single biggest success factor in a digital transformation journey, and **most organisations are ill prepared** due to many islands of disconnected data that is of unknown and/or poor quality.

Introduction to Data Migration

Sadly, Data only gets into the spotlight when an organisation Is faced with a large system upgrade. Unfortunately, this is typically in a negative light when it is exposed as a major risk to a program with a major cost attached to remediate the data and mitigate the risk. This Is a symptom of being under-valued in a traditional business model that has survived on paper-based systems and manual data entry which is not sustainable in the digital age.

> ### *Takeaway #1*
> *Data will typically be exposed as a major risk and cost in large scale digital transformations due to unknown and/or poor quality.*

Figure 1: The Data Migration Spotlight
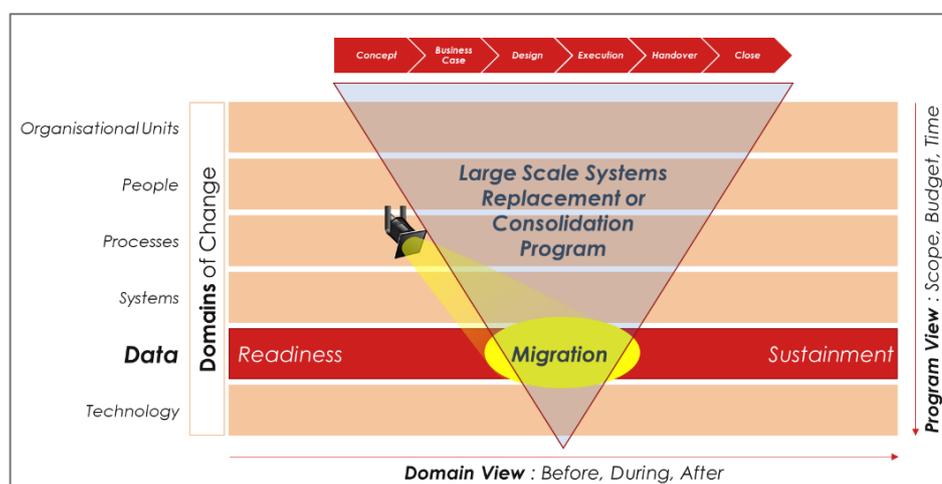
As a "domain", Data should and can be managed before, during and after a major program. This strengthens organisations overall digital capability and mitigates future risks and costs by ensuring data remains evergreen.

Data Migration can best be explained as three distinct phases

Pre-Migration:	**Data Profiling & Remediation** *(White Paper #1)*
During Migration:	**Data Standardisation & Loading** *(White Paper #2)*
Post-Migration:	**Data Reconciling & Archiving** *(White Paper #3)*

This paper focusses on the Pre-Migration phase.

Figure 2 : Data Migration Phases

The Pre-Migration phase is preparing your organisations data for migration. Establishing Data Owners who can then contribute to and approve Data Quality Management (DQM) objectives and guiding principles will define the profiling and remediation requirements of the pending data migration.

Data Profiling is the systematic analysis of the source data based on the requirements given. The desired outcome of the Profiling activity is to provide a correct and complete model for the target solution. Once data has been assessed, the results of the Profiling will drive the Data Remediation by determining the actions needed to cleanse and prepare the source data.

Data Remediation can involve correcting duplicate, incomplete, inaccurate, and corrupt records in existing systems. There may also be a requirement for the standardisation and harmonisation of records to align with the requirements of the new system. There are six dimensions of data quality as outlined in Table 1: Six Dimensions of Data Quality shown **below**.

When and Where to Start

A **common practice** and misconception in traditional, less digitally mature business is that a program, together with some technology, will fix whatever issues are encountered, and once fixed It will be fixed for good. Two things happen as a result:

1. The program risks and costs increase significantly due to a lack of data Readiness; and
2. The program will, by nature, only focus on doing the minimum work required for the program to achieve its objectives, often described as a new system going live. This approach most often does not include ensuring the business has a Sustaining capability and as a result, data will sadly fall into a state of disrepair over time, and the benefits of the transformation will not be sustained.

Figure 3: Common vs. Best Practice Approach to Data Migration

A **better/best practice** Is to acknowledge that this Is a business issue first and foremost and that the business must take ownership and accountability of its data as a strategic asset.

For each Master Data object (Finance, Supplier, People etc), a senior business person, typically from the relevant business function, should be assigned the role and responsibilities of a Data Owner *(see Appendix A for a guide on roles & responsibilities).*

> *Takeaway #2*
> *Strong Data Ownership and Data Governance Is Critical*

Data Owners, along with the organisations **Process Owners** should meet regularly in an Enterprise Process and Data Owners forum to discuss, review and agree on performance, quality and remediation activities needed to continually improve business performance.

Figure 4: Example Enterprise Process and Data Owners

In-Program vs Pre-Program Data Remediation

As mentioned above, a common misconception is that program commencement is needed to kick-off Data Profiling & Remediating.

Per **Figure 3: Common vs. Best Practice Approach to Data Migration**, a better practice Is for Data Owners to commence Data Profiling as soon as practical to understand the baseline quality of their data.

> ### Takeaway #3
> *Commence Data Profiling & Data Remediating as early as possible.*

Data Quality can be improved ahead of a Data Migration program which reduces time, effort, cost and risk to the actual program, and further strengthens the sustaining capability to maintain data quality beyond the program completion.

Initial Data Quality efforts should focus on the top 3 Dimensions of Data Quality Dimensions shown here in **Table 1: Six Dimensions of Data Quality.**

Business	**Duplication**	Duplicate records may exist. (e.g. specific master data may exist multiples times with each instance having a variation to the original master data naming convention). Any duplicates are to be deleted and relationships amended to link the surviving row.
	Redundancy	Data that is no longer current. These should be identified in source systems and corrected accordingly. (e.g. Active vendors without an invoice in the last 14 months).
Program	**Standardisation**	The data to be migrated will need to conform to an approved or conventional standard (e.g. Master Data Standard). Cross-validation of datasets across table structures against the agreed standards will need to be monitored and rectified as appropriate until go-live.
	Incorrect	The data has the incorrect business value (e.g. Entitlement amount is incorrect, Bank Account name is incorrect, Addresses inclusive of post codes are incorrect, field values are not aligned to their original planned usage patterns)
	Integrity	Relationships are not maintained correctly, such as orphaned records (e.g. an Account Balance for an Account that no longer exists). Tables and application functional areas must be maintained (e.g. Organisational Structure, Accounts, Accounts, Reports to etc).
	Completeness	Is a measure of data content quality expresses as a percentage of the columns or fields of a table or file that should have values in them, or fields that need to be left blank and have values in them. (e.g. First Names not in Preferred Name column).

Table 1: Six Dimensions of Data Quality

Data Profiling and Data Remediating Takeaways

1. **Data will typically be exposed as a major risk and cost** in large scale digital transformations due to unknown and/or poor quality. If you don't know the quality of your data today, this will likely be true for you.

2. **Improve your Readiness** by:

   a. **establishing strong Data Ownership and Data Governance immediately**. These roles are needed now and will be required into your digital future. Start now and begin to develop that capability.

   b. **ensuring the Data Owners define Data Quality objectives as their priority task**. These objectives are used for profiling the data and baselining your current situation. Without this, your organisation and any pending programs are flying blind.

   c. **starting Data Remediation as soon as possible.** This Is the most common pitfall, where organisations leave It to the 'program' to fix the problem. Data Owners can, and should, resolve duplicate data and redundant data in existing systems. There Is no dependency on new systems for this to occur, and this exercise alone helps develop a critical enduring capability.

We hope that these guidelines help you in some way with your digital transformation journey in this critical, yet often overlooked area of Data Migration.

The next white paper in this series will focus on the next phase - **Data Loading**.
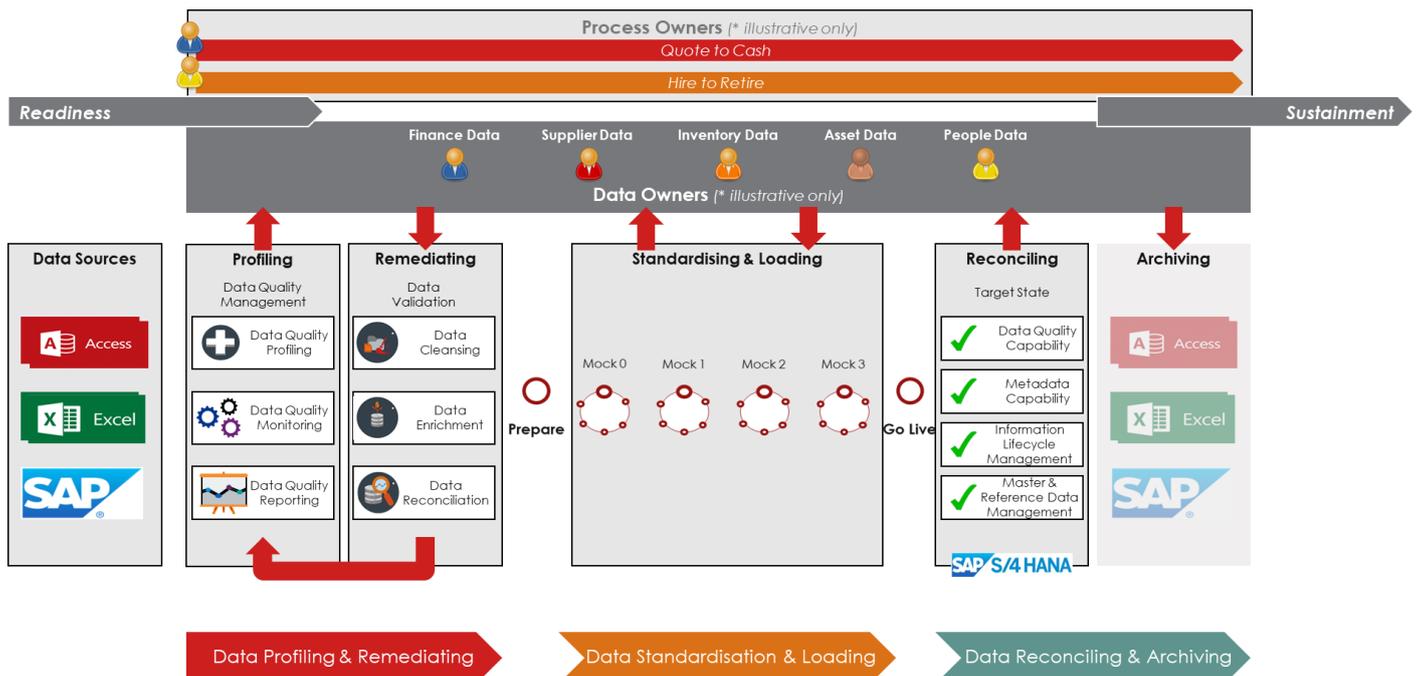
ooOoo

Figure 5: This Series will explore Data Owners journey throughout Data Migration

Appendix A : Data Owners and Data Steward Roles & Responsibilities Guide

Data Owners	Data Stewards
Business Role • Overall responsibility, ownership, and authority for a set of business data, usually in their area of business expertise. • Typically, directly affected by the data's accuracy, integrity, and timeliness in their day to day activities. • Approves/endorses the data migration results based on the reviews performed by the Data Stewards.	**Business Role** • The person that works with or uses the data on a day to day basis. They are involved in creating and maintaining the data (sometimes with the assistance of Data Custodians). • Is affected the most by poor quality data but is also best placed to resolve data quality issues.

Key Responsibilities
• Resourcing
  o Nominates and/or endorses Data Stewards
• Governance (*)
  o Sign off Data Reconciliation Plan
  o Sign off Trial Data Migrations
  o Approve mitigation plan for Data Quality issues unlikely to be cleansed by go-live
  o Sign off Final Data Migration
  o Maintain Data Quality post Go Live
	Key Responsibilities
	• Defining Target Data Set
	  o Participate in development of Data Migration Requirements (including Data Extraction and Mapping Rules)
	• Data Quality / Cleansing Activities
	  o Participate in Data Quality Reviews led by Data Migration Team
	  o Assign actions from Data Quality Reviews
	  o Perform data cleansing and data collection
	  o Profile and analyse preload and post-load data
	• Data Reconciliation
	  o Participate in Data Reconciliation Planning
	  o Contribute to and validate the Data Reconciliation Plan
	  o Perform data reconciliation for the trial and Go-Live data loads
	  o Send endorsement to Data Owners confirming that the data has been cleansed, validated, and reconciled with support from the Data Migration Team