



# **Data Profiling, Cleansing, Transformation & Migration (PCTM) Methodology**

## TABLE OF CONTENTS

1 INTRODUCTION.....	3
2 DATA MANAGEMENT APPROACH.....	3
2.1 Overview.....	3
2.2 Stages of The Data Management Process .....	3
3 PROJECT INITIATION AND PLANNING .....	4
4 DATA PROFILING STAGE.....	4
4.1 Data Profiling stage Life Cycle.....	4
5 DATA CLEANSING STAGE.....	5
5.1 Data Cleansing stage life cycle .....	6
6 DATA TRANSFORMATION STAGE.....	7
6.1 Data Transformation stage Life cycle .....	7
7 DATA MIGRATION STAGE.....	9
7.1 Data Migration Stage Life cycle.....	9
8 DATA VALIDATION AND VERIFICATION.....	11
8.1 Validation & Verification stage Life cycle .....	11
9 APPENDIXES.....	13
9.1 Appendix A – PCTM – Methodology.....	13
9.2 Appendix B – PCTM - Process .....	14

## 1 INTRODUCTION

This document describes the Datalynx PCTM Methodology for Profiling, Cleansing, Transforming, and Migrating data from source database(s) into one or more destination databases.

The scope of this document is to outline the following:

- Data management approach
- Data Profiling stage
- Data Cleansing stage
- Data Transformation stage
- Data Migration stage
- Data validation and verification

## 2 DATA MANAGEMENT APPROACH

### 2.1 OVERVIEW

The PCTM Methodology defines a proven, systematic approach for delivering data-related solutions. The approach comprises iterative-incremental and parallel development activities that support delivery of project elements over a number of business analysis and technical processing cycles.

Iterative cycles are a key component of each of the major project stages (profiling, cleansing, transformation and migration) that help to maximise agility and the ability to respond to evolving needs. This allows the project team to utilise the results and feedback of each cycle to address changes in business requirements and incorporate those outcomes into subsequent activities.

The Datalynx PCTM Methodology leads to improved quality of project deliverables, as well as enhanced productivity and contributes significantly to successful project delivery.

### 2.2 STAGES OF THE DATA MANAGEMENT PROCESS

The end-to-end data management process specified by the Methodology consists of six clearly defined stages, which are further illustrated in Appendixes A – B of this document:

1. Project Initiation and Planning.
2. Data profiling.
3. Data cleansing.
4. Data transformation.
5. Data migration.
6. Data validation.

### 3 PROJECT INITIATION AND PLANNING

The Project Initiation and Planning stage is designed to specify the datasets and define the activities that are in scope for the project, as well as project deliverables.

Note: This document focuses on data management related activities only and standard project tasks such as establishing governance, identifying stakeholders and support groups and specifying reporting requirements are not defined.

### 4 DATA PROFILING STAGE

The Data Profiling stage is designed to identify quality issues in the source dataset by utilising the following techniques:

- Specifying and evaluating common profiling rules for individual data elements to identify any common data issues and to ensure source data compatibility with the target system.

Common profiling analysis includes field type validation, checking for blank / null values and value frequency analysis of specific fields.

- Specifying and evaluating custom business rules on individual or multi-field data elements to identify business-specific data issues, and to ensure data compatibility with the target system.

Custom profiling analysis includes identifying duplicate records, broken data relationships (orphan records) and ensuring data values for specific fields are within business-defined ranges and sets.

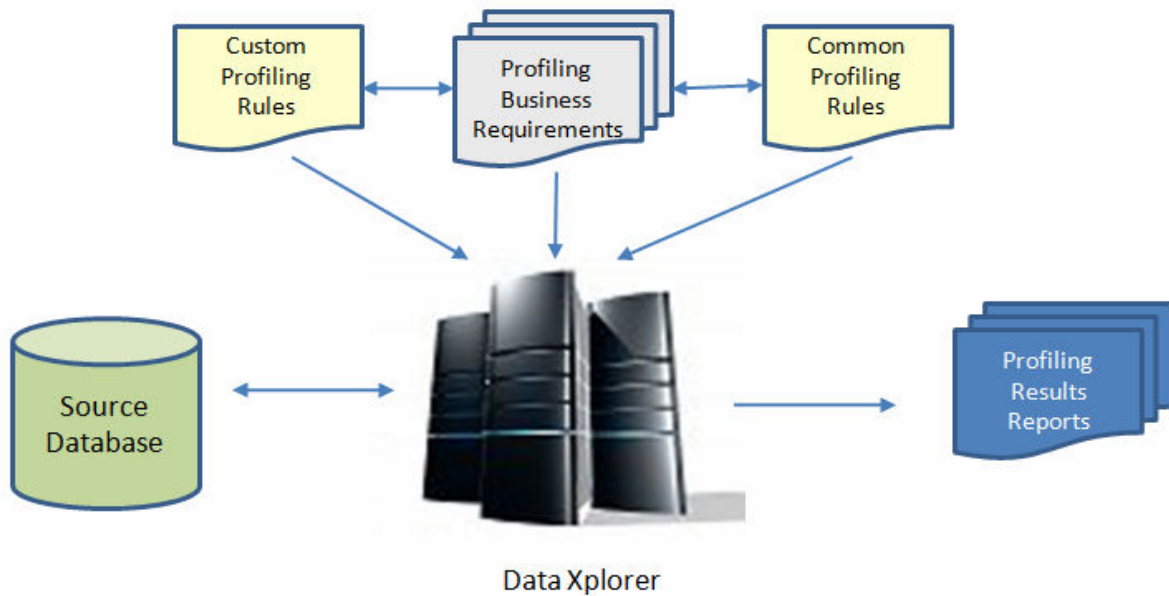
- Producing issue reports and statistical information that show the results of the various checks against the source dataset.

#### 4.1 DATA PROFILING STAGE LIFE CYCLE

Data Profiling inputs and outputs:

Input	Data Management Tools	Output
<ul style="list-style-type: none"> <li>• Source (Staging) database</li> <li>• Common profiling rules</li> <li>• Custom profiling rules</li> <li>• Profiling business requirements</li> </ul>	<p>Data Xplorer</p>	<p>Data profiling results reports</p>

The following diagram illustrates the relevant components and interactions in the Data Profiling Stage:



## 5 DATA CLEANSING STAGE

The data quality issues that were identified during the Data Profiling Stage are documented in a Summary Findings Report and a detailed Profiling Results Report. These findings provide the basis for specifying the data cleansing business rules that will be applied during the Data Cleansing Stage. Similar to the Profiling stage, cleansing rules are defined as common rules (applicable to all nominated fields of a particular type) and custom rules, to address specific business issues.

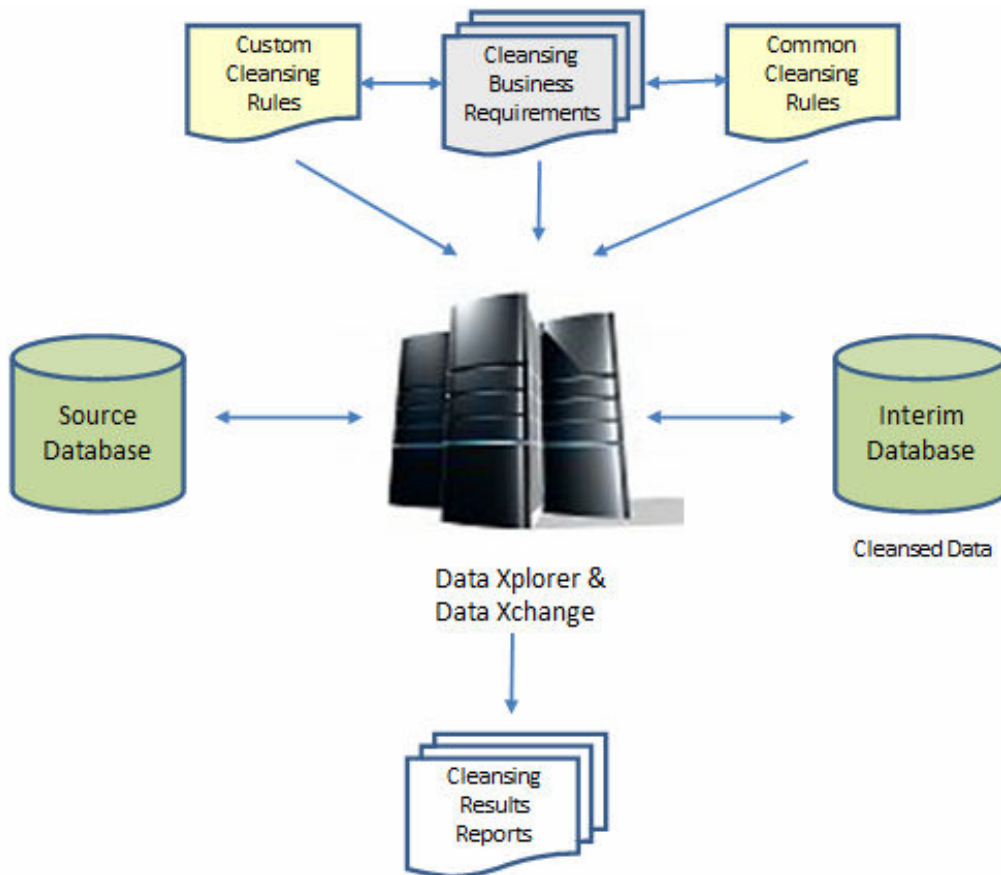
Data cleansing is not executed against the source data as it is necessary to keep the original data unaltered for audit and reporting purposes. An “interim” database with an identical structure to the source database is created and a copy of the source data is migrated across as part of the cleansing process. Where there are source tables for which no cleansing is to be applied, their data is copied across as is.

### 5.1 DATA CLEANSING STAGE LIFE CYCLE

Data Cleansing Inputs and outputs:

Input	Data Management Tools	Output
<ul style="list-style-type: none"> <li>• Source (Staging) database</li> <li>• Common cleansing rules</li> <li>• Custom cleansing rules</li> <li>• Data cleansing business requirements</li> </ul>	<ul style="list-style-type: none"> <li>• Data Xchange</li> <li>• Data Xplorer</li> </ul>	<ul style="list-style-type: none"> <li>• Detailed data cleansing reports &amp; recommendations</li> <li>• Cleansed (Interim) database</li> </ul>

The following diagram illustrates the relevant components and interactions in the Data Cleansing Stage:



## 6 DATA TRANSFORMATION STAGE

The Data Transformation Stage takes the cleansed data stored in the Interim database during the preceding Data Cleansing Stage and updates the structure and content to a standard format as specified in the Transformation business rules.

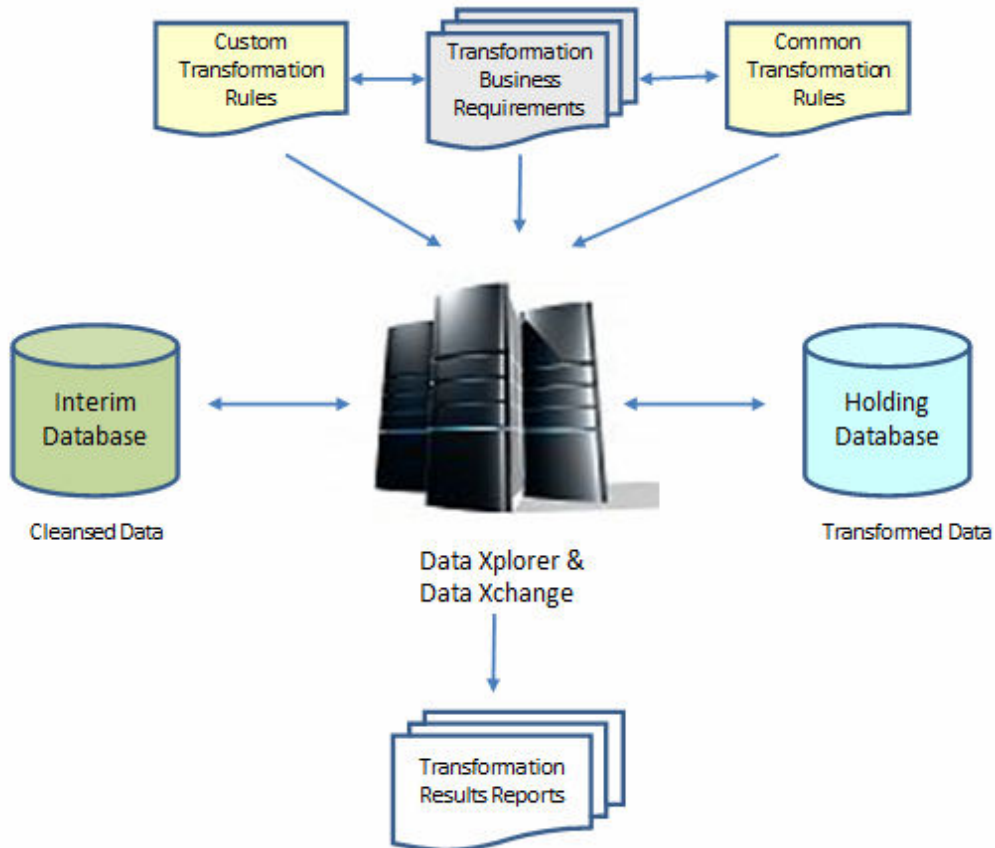
Transformation typically includes activities such as data consolidation, standardisation and segregation of data field components, as well as creation of derived fields. During the Data Transformation Stage, data is transformed from the copy of the source structure in the Interim database, to a new predefined structure in a Holding database.

### 6.1 DATA TRANSFORMATION STAGE LIFE CYCLE

Data transformation inputs and outputs:

Input	Data Management Tools	Output
<ul style="list-style-type: none"> <li>• Cleansed (Interim) Database</li> <li>• Common transformation rules</li> <li>• Custom transformation rules</li> <li>• Data transformation business requirements</li> </ul>	<ul style="list-style-type: none"> <li>• Data Xchange</li> <li>• Data Xplorer</li> </ul>	<ul style="list-style-type: none"> <li>• Detailed data transformation Reports &amp; recommendations</li> <li>• Cleansed/Transformed (Holding) database</li> </ul>

The following diagram illustrates the relevant components and interactions in the Data Transformation Stage:



## 7 DATA MIGRATION STAGE

The Data Migration Stage is where the transfer of cleansed, transformed data is undertaken from the Holding database to the new system’s database(s). During this stage there can be further changes to the format of data in the standard database structures of the Holding database to align with the database structure of the target system.

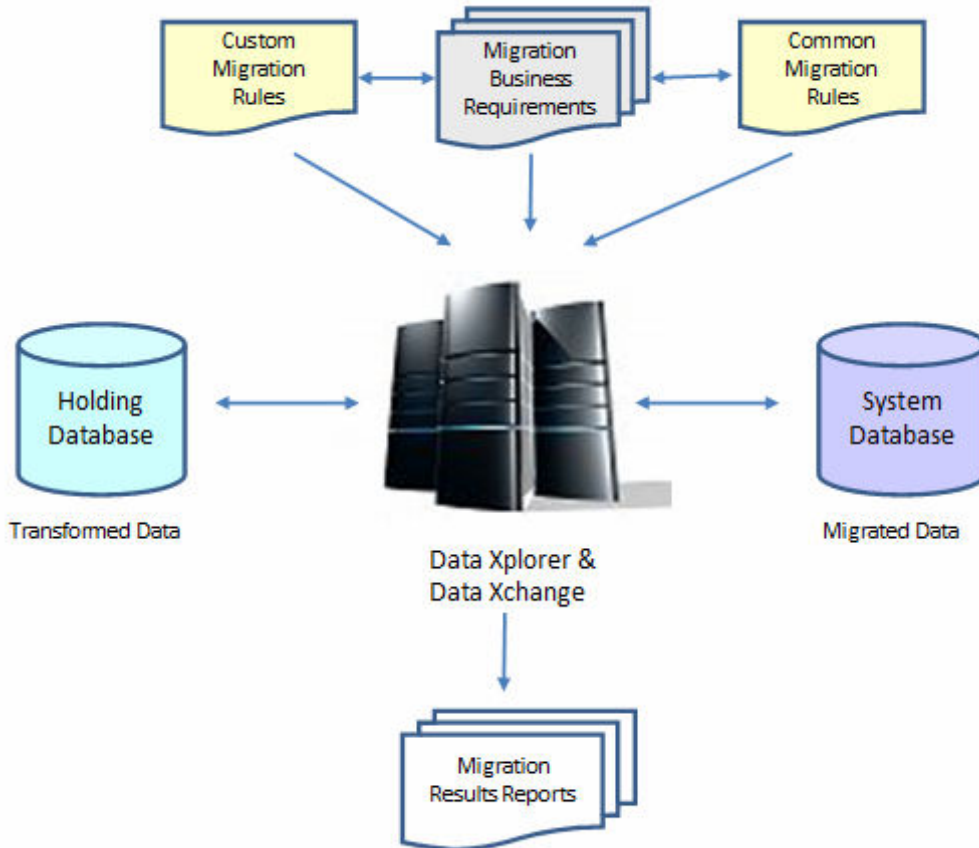
An additional option that is considered as part of the Data Migration is the adoption of a two-phased approach whereby historical data (or data up to a particular point in time) is migrated in the first phase and newer / in progress transactions are migrated subsequently. This helps to minimise system downtime during cutover and reduces the impact on business activities.

### 7.1 DATA MIGRATION STAGE LIFE CYCLE

Data Migration Stage inputs and outputs:

Input	Data Management Tools	Output
<ul style="list-style-type: none"> <li>• Cleansed/Transformed (Holding) database</li> <li>• Common migration rules</li> <li>• Custom migration rules</li> <li>• Data migration business requirements</li> </ul>	<ul style="list-style-type: none"> <li>• Data Xchange</li> <li>• Data Xplorer</li> </ul>	<ul style="list-style-type: none"> <li>• Detailed data migration reports &amp; recommendations</li> <li>• Final (system) database</li> </ul>

The following diagram illustrates the relevant components and interactions in the Data Migration Stage:



## 8 DATA VALIDATION AND VERIFICATION

Data validation and verification is undertaken at the end of each cycle in the Profiling, Cleansing and Transformation stages, as well as being the final activity post the Migration.

Following the final migration cycle, a complete set of reports is produced that provides end-to-end data verification and validation results and serves as an audit trail of activities that have been undertaken.

### 8.1 VALIDATION & VERIFICATION STAGE LIFE CYCLE

Data Validation and Verification inputs and outputs:

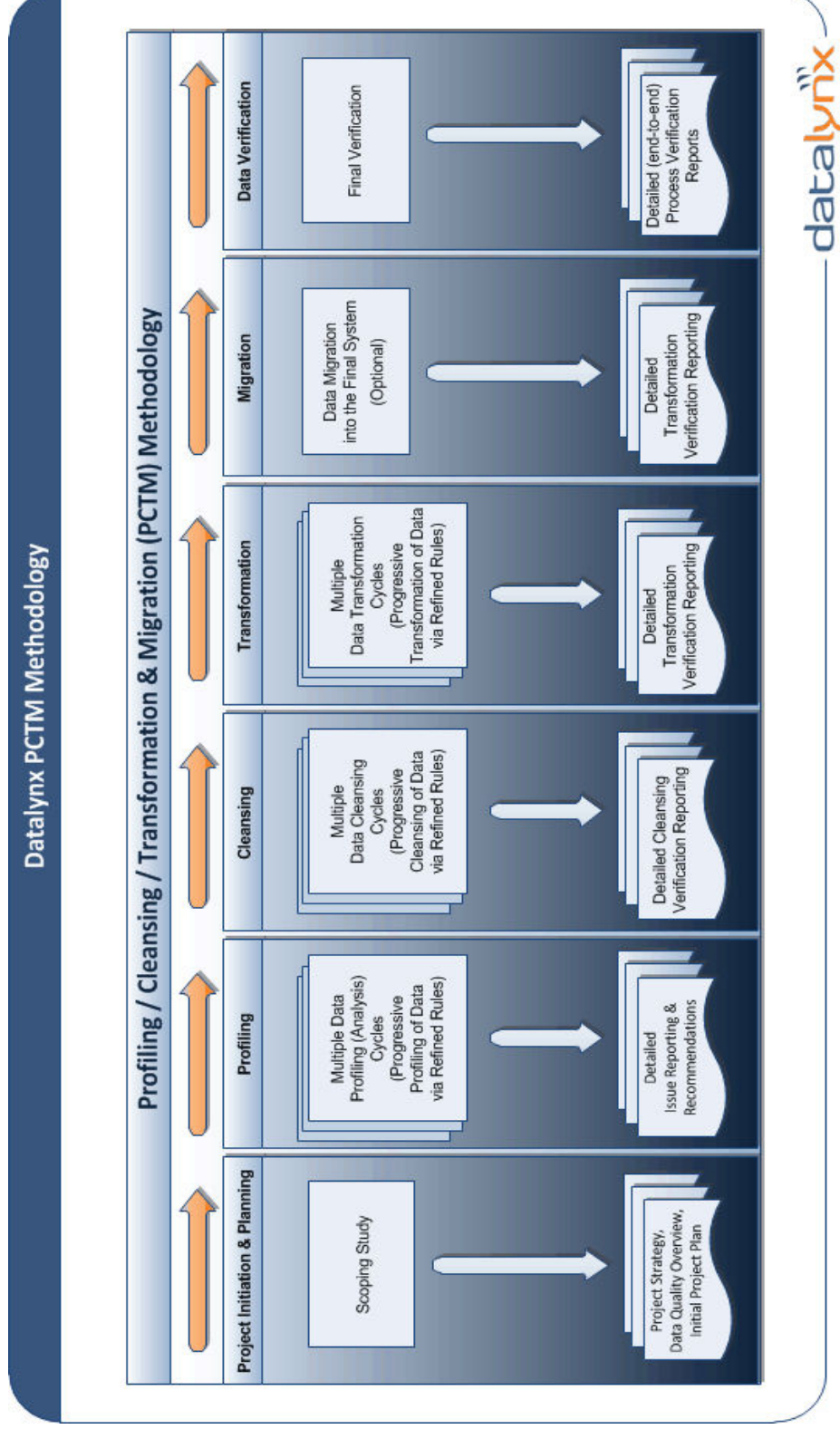
Input	Data Management Tools	Output
<ul style="list-style-type: none"> <li>• Source Database (staging)</li> <li>• Cleansed Database (Interim)</li> <li>• Transformed Database (Holding)</li> <li>• Migrated Database (System)</li> </ul>	<ul style="list-style-type: none"> <li>• Data Xchange</li> <li>• Data Xplorer</li> </ul>	<ul style="list-style-type: none"> <li>• Data profiling reports</li> <li>• Data cleansing reports</li> <li>• Summary report</li> <li>• Recommendations</li> <li>• Database metadata and record counts at the end of each stage of the process.</li> </ul>



## 9 APPENDICES

### 9.1 APPENDIX A – PCTM – METHODOLOGY

The following diagram illustrates the key activities and deliverables defined by the PCTM Methodology:



## 9.2 APPENDIX B – PCTM - PROCESS

The following diagram illustrates the practical application of the Datalynx PCTM process:

