Datalynx Project Delivery Methodology and PCTM Methodology

For

Legacy Data Cleansing & Migration

Title:	Datalynx PCTM Methodology for Migrating Legacy Data
Revision:	V1.4

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

Table of Contents

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

1 INTRODUCTION

Legacy data migration is a critical component of business transformation and IT system modernisation initiatives that is commonly a source of significant project risk.

Traditional approaches that address data migration as a final step prior to implementing a new system face a high probability of failure and at the very least result in timeframe and cost blowouts.

Datalynx has developed and utilises the agile PCTM Methodology for data migration. This approach takes data cleansing and migration off a project's critical path and has a proven track record of success on complex projects, delivering required outcomes when other models have failed.

This document describes the Datalynx methodology for cleansing and migrating data.

Specifically it focuses on:

- Datalynx's Data Migration Project Methodology
- Data Migration Project stages and activities:
    - Data Profiling stage
    - Data Cleansing stage (as required)
    - Data Transformation stage (as required)
    - Data Migration stage
    - Data Validation and Verification stage
- The agile PCTM Methodology for guiding data cleansing / migration activities during the project and managing business risk.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2 DATA CLEANSING / MIGRATION PROJECT APPROACH

2.1 OVERVIEW

The Datalynx PCTM Methodology defines a proven, systematic approach for delivering data cleansing and migration solutions. The approach comprises iterative, incremental and parallel solution development activities that support delivery of project objectives over a number of cycles per project stage.

Iterative cycles are a key element of each of the major project stages (profiling, cleansing, transformation and migration) that help to maximise agility and the ability to respond to evolving business needs. This allows the project team to refine processing configuration to address changes in business requirements and incorporate the outcomes of earlier cycles.

The Datalynx PCTM methodology contributes to improved quality of project deliverables, enhanced productivity and successful project delivery.

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.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.2 PROJECT STAGES

The Datalynx Data Migration Project delivery methodology consists of six discrete stages, specifically:

1. Project Initiation and Planning
2. Data Profiling
3. Data Cleansing
4. Data Transformation
5. Data Migration
6. Data Validation

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.3 DATA MIGRATION PROJECT INITIATION AND PLANNING

2.3.1 OVERVIEW

The Project Initiation and Planning stage focuses on project scoping and associated planning activities. Project scoping provides early notice of any challenges or issues that need ot be factored into the planning process.

During this stage the following tasks are completed:

ID	Activity
1	Specify the business areas / system modules which are in scope for the project.
2	Identify the source databases / datasets which are in scope.
3	Review the target data model (and its level of completeness, if not finalised at this stage) and any available documentation / supporting material relating to the data structures of the legacy and target systems.
4	Determine and agree the number of profiling cycles which will be undertaken (at least one profiling cycle is required). Note: The number of profiling cycles ultimately required is a factor of the level of business knowledge of data issues and the overall quality of the source data.
5	Agree the number of cleansing cycles which are to be undertaken (at least two cleansing cycles are required).
6	Agree the number of transformation cycles which are in scope (at least two transformation cycles are required).
7	Agree the number of migration cycles which are in scope (at least two migration cycles are required).
8	Develop a detailed project management plan and schedule.

Note: Datalynx also offers stand-alone Scoping Analysis services to support project sizing and estimation, business case development and overall data migration risk management.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.4 DATA PROFILING STAGE

2.4.1 OVERVIEW

The Data Profiling stage is designed to identify legacy data quality issues, as well as data that does not comply with the new system's configuration rules, by utilising the following techniques:

- Specifying and evaluating generic 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, data completeness checking and a range of statistical analysis measures.

- Specifying and evaluating custom business rules against 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, verifying data integrity 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 legacy data.

2.4.2 INITIAL DATA PROFILING CYCLE

The initial data profiling cycle consists of the following activities:

ID	Activity
1	Register the source databases in the Datalynx Data Xplorer system
2	Specify and document data profiling (analysis) business requirements
3	Design the generic and custom profiling functions required
4	Create and configure the data profiling templates in Data Xplorer
5	Execute data profiling processing
6	Generate profiling reports
7	Perform post profiling data analysis

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

ID	Activity
8	Present the profiling results and the data analysis findings
9	Review the data profiling business requirements and business rules with business representatives and SMEs .

2.4.3 SUBSEQUENT DATA PROFILING CYCLE(S)

The subsequent data profiling cycle consists of the following activities:

ID	Activity
1	Take a new copy of the legacy data from production (only necessary if there have been significant data updates since the last copy) and re-register the legacy databases in the Datalynx Data Xplorer system.
2	Refine and update data profiling business requirements
3	Refine and update the common and custom profiling functions
4	Update the configuration of the Data Xplorer system
5	Execute the updated data profiling jobs and generate profiling results reports
6	Perform post profiling data analysis
7	Present the profiling results and the data analysis findings
8	Review the data profiling business requirements and business rules to determine if further data profiling cycles are required

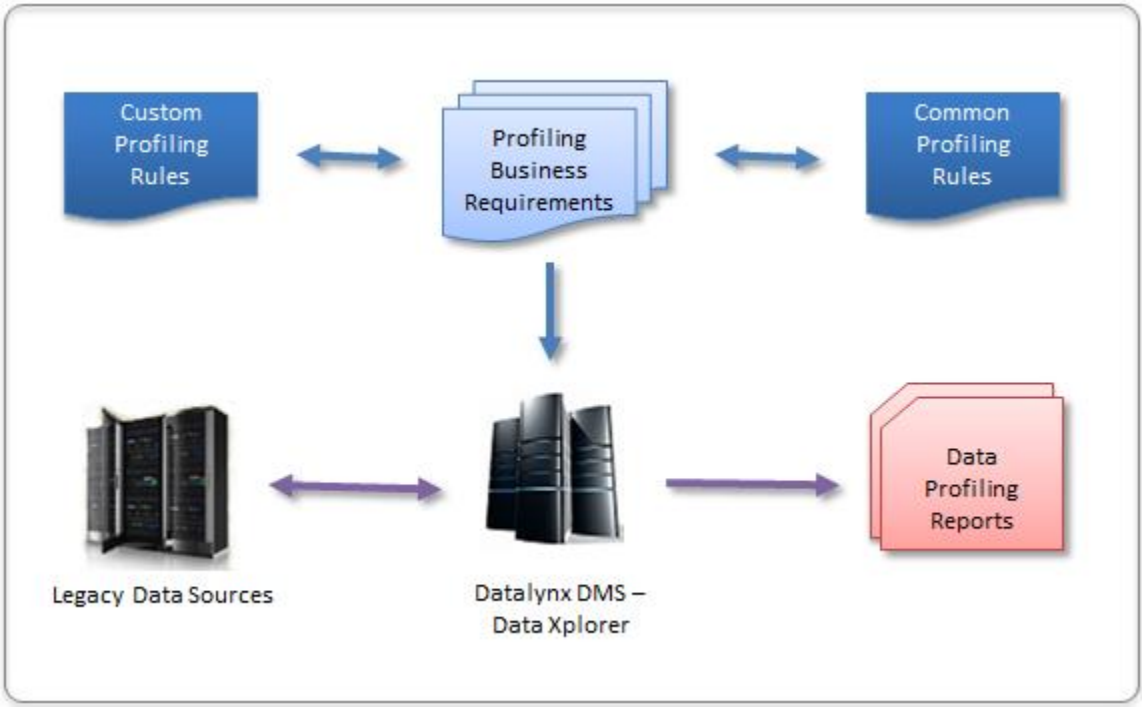
© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.4.4 DATA PROFILING STAGE LIFECYCLE

Data Profiling inputs and deliverables:

Input	Data Management System	Deliverables
• Source (legacy) databases • Common profiling rules • Custom profiling rules • Profiling business requirements	Datalynx Data Xplorer	Data profiling results reports

The following diagram illustrates the Data Profiling Stage life cycle:



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.5 DATA CLEANSING STAGE

It is highly recommended that data cleansing is undertaken as a component of a data migration initiative. Poor source data quality and misalignment between legacy data and the requirements of the new (destination) system are common causes of data migration problems.

2.5.1 OVERVIEW

The data quality issues that were identified during the Data Profiling Stage are analysed and documented. 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.

2.5.2 INITIAL DATA CLEANSING CYCLE

The initial data cleansing cycle consists of the following activities:

ID	Activity
1	Create an "Interim" database to serve as the destination for the cleansed data from each legacy database.
2	Specify cleansing business requirements.
3	Specify the common and custom cleansing rules.
4	Establish connections to the source (legacy) databases and the destination (Interim) databases using the Datalynx Data Xchange system.
5	Create and configure data cleansing templates in Data Xchange.
6	Execute data cleansing processing and generate the data cleansing reports.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

ID	Activity
7	Perform post cleansing data analysis that includes: • Inspection of the data cleansing reports • Review and validation of the data cleansing results
8	Present the cleansing results and data analysis findings.
9	Review the data cleansing business requirements and business rules with business representatives and SMEs.

2.5.3 SUBSEQUENT DATA CLEANSING CYCLE(S)

The subsequent data cleansing cycle(s) consist of the following activities:

ID	Activity
1	Based on the results of the previous cleansing cycle, refine cleansing business requirements as required.
2	Refine the common and custom cleansing rules.
3	Update the configuration of the data cleansing templates in Data Xchange.
4	Execute data cleansing processing and generate the data cleansing reports.
5	Perform post cleansing data analysis that includes: • Inspection of the data cleansing reports • Review and validation of the data cleansing results
6	Present the cleansing results and data analysis findings.
7	Review the data cleansing business requirements and business rules to determine if further data cleansing cycles are required.

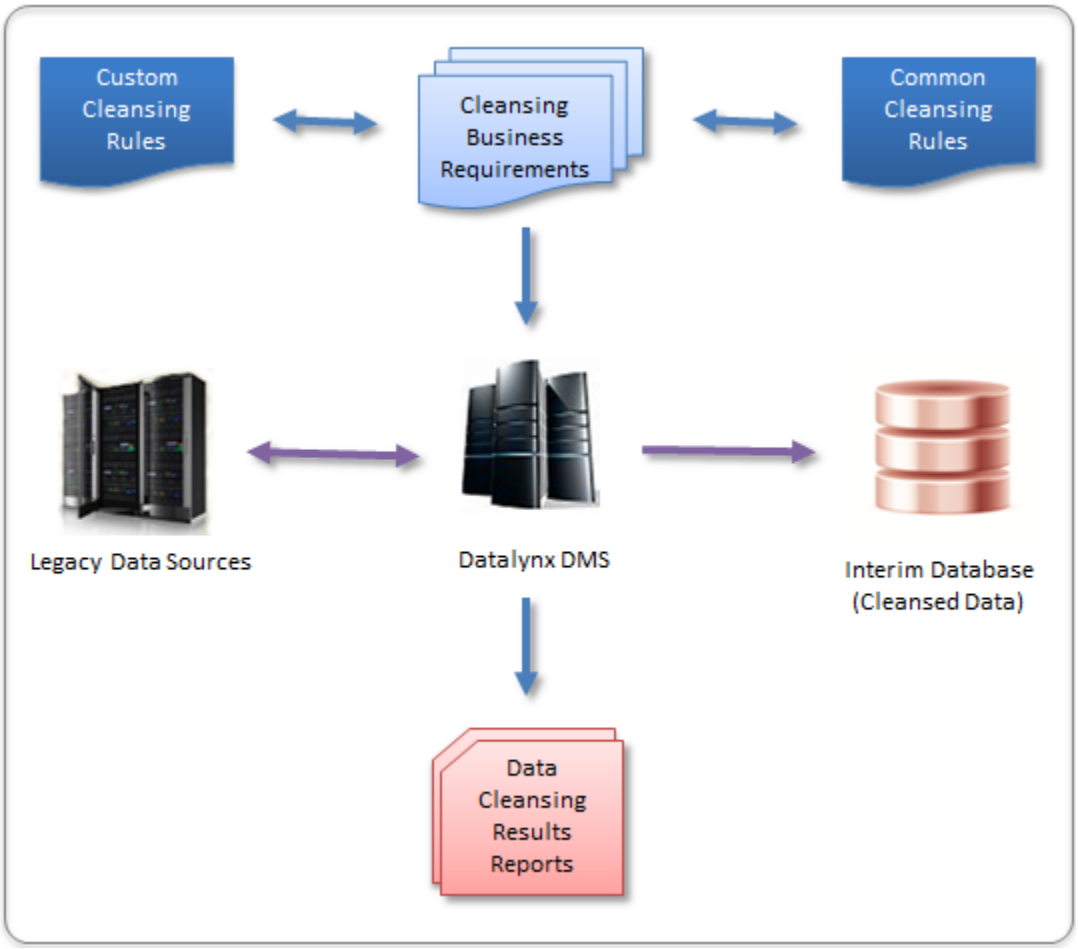
© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.5.4 DATA CLEANSING STAGE LIFECYCLE

Data Cleansing inputs and deliverables:

Input	Data Management Systems	Deliverables
• Source (legacy) databases • Common cleansing rules • Custom cleansing rules • Data cleansing business requirements	• Datalynx Data Xchange • Datalynx Data Xplorer	• Detailed data cleansing reports & recommendations • Cleansed (Interim) database

The following diagram illustrates the Data Cleansing Stage life cycle:



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.6 DATA TRANSFORMATION STAGE

2.6.1 OVERVIEW

The Data Transformation Stage takes the cleansed data stored in the Interim database during the 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 legacy structure in the Interim database, to a new predefined structure in a Holding database.

2.6.2 INITIAL DATA TRANSFORMATION CYCLE

The initial data transformation cycle consists of the following activities:

ID	Activity
1	Define and create the destination (Holding) database that will hold the transformed data.
2	Specify transformation business requirements.
3	Specify the common and custom transformation rules.
4	Establish connections to the source (Interim) database and the destination (Holding) database using the Datalynx Data Xchange system.
5	Create and configure the data transformation templates in Data Xchange.
6	Execute the data transformation jobs and generate the data transformation reports.
7	Perform post transformation data analysis that includes: • Inspection of the data transformation reports • Review and validation of the data transformation results
8	Present the transformation results and data analysis findings.
9	Review the data transformation business requirements and business rules with business representatives and SMEs.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.6.3 SUBSEQUENT DATA TRANSFORMATION CYCLE(S)

The subsequent data transformation cycles comprise the following activities:

ID	Activity
1	Refine the transformation business requirements.
2	Refine the common and custom transformation rules.
3	Update the configuration of the data transformation templates in Data Xchange.
4	Execute the data transformation jobs and generate the data transformation reports.
5	Perform post transformation data analysis that includes: • Inspection of the data transformation reports • Review and validation of the data transformation results
6	Present the transformation results and data analysis findings.
7	Review the data transformation business requirements and business rules to determine if further data transformation cycles are required.

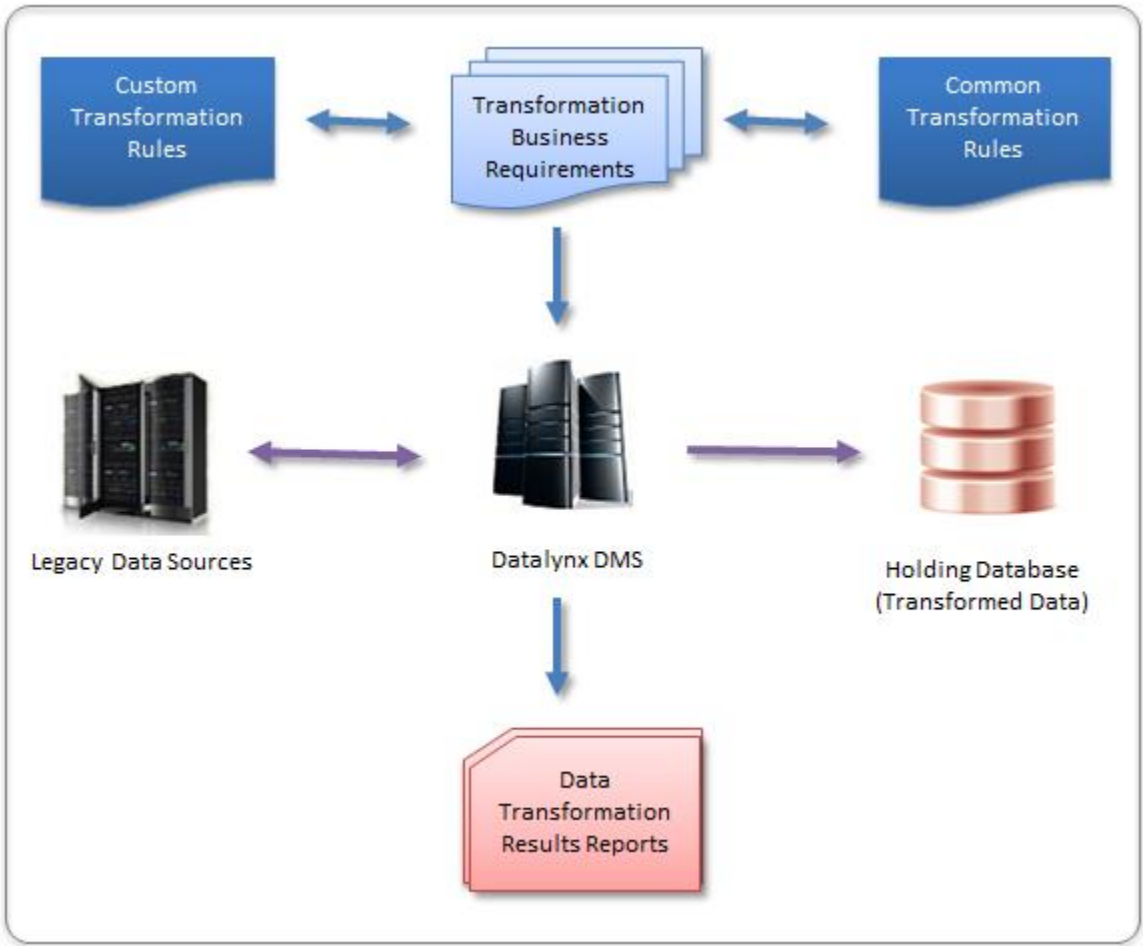
© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.6.4 DATA TRANSFORMATION STAGE LIFE CYCLE

Data transformation inputs and deliverables:

Input	Data Management Systems	Deliverables
• Cleansed (Interim) Database • Common transformation rules • Custom transformation rules • Data transformation business requirements	• Datalynx Data Xchange • Datalynx Data Xplorer	• Detailed data transformation Reports & recommendations • Cleansed/Transformed (Holding) database

The following diagram illustrates the Data Transformation Stage life cycle:



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.7 DATA MIGRATION STAGE

2.7.1 OVERVIEW

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

An option that is commonly utilised for the Data Migration is two-phased approach whereby historical data (or data up to a 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.

2.7.2 DATA MIGRATION CYCLE

The initial data migration cycle consists of the following activities:

ID	Activity
1	Define the source (Holding) database.
2	Define the destination (System) database.
3	Specify migration business requirements.
4	Specify the common and custom data migration rules.
5	Establish connections to the source (Holding) database and the destination (System) database using the "Data Xchange" system.
6	Create and configure the data migration templates in Data Xchange.
7	Execute the data migration jobs.
8	Generate the data migration reports.
9	Perform post migration data analysis that includes: • Inspection of the data migration reports • Review and validation of the data migration results
10	Present the data migration results and data analysis findings.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

ID	Activity
11	Review the data migration business requirements and business rules with business representatives and SMEs to determine if additional cycles are required.

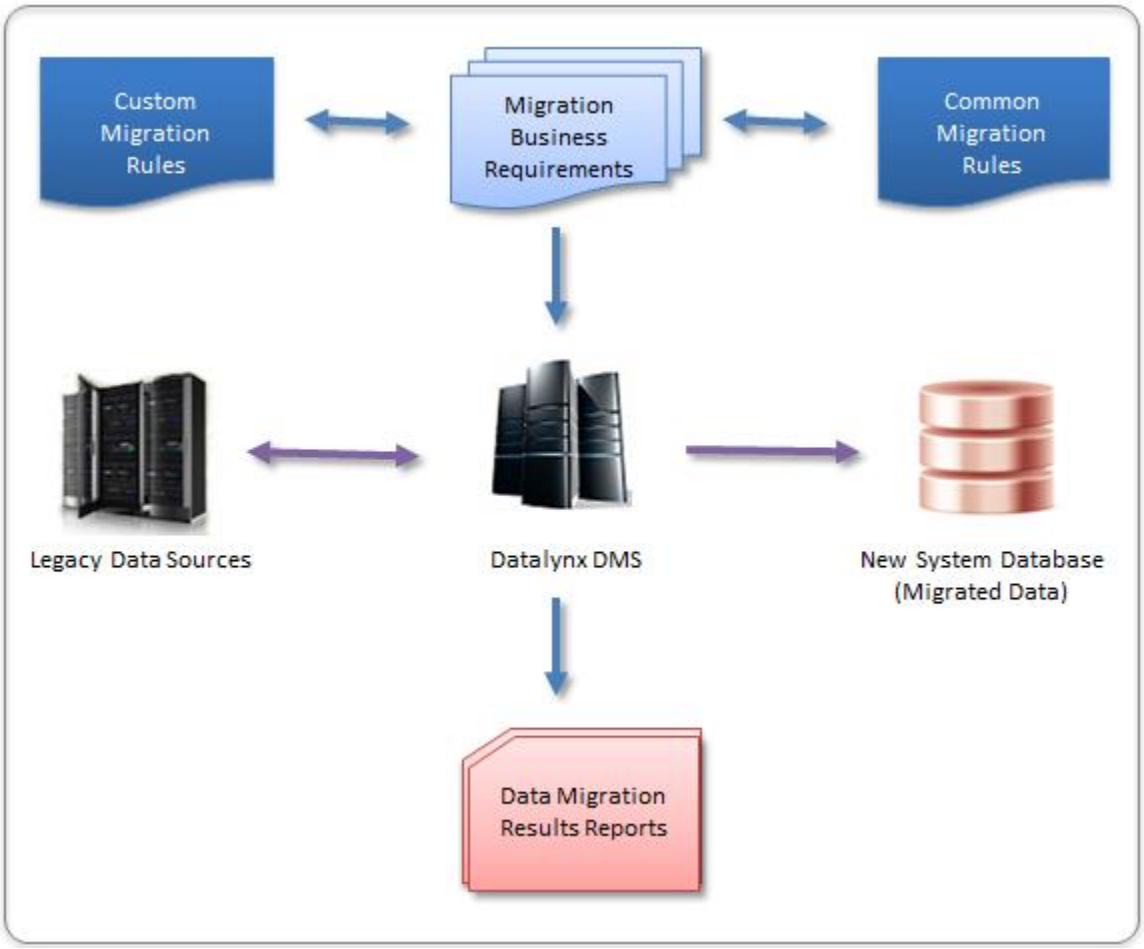
2.7.3 DATA MIGRATION STAGE LIFE CYCLE

Data Migration Stage inputs and deliverables:

Input	Data Management Systems	Deliverables
• Cleansed/Transformed (Holding) database • Common migration rules • Custom migration rules • Data migration business requirements	• Datalynx Data Xchange • Datalynx Data Xplorer	• Detailed data migration results reports • Final (system) database

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

The following diagram illustrates the Data Migration Stage life cycle:



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.8 DATA VALIDATION AND VERIFICATION STAGE

2.8.1 OVERVIEW

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 stage.

Following the final migration cycle, a complete set of reports is produced that provides end-to-end data verification and validation results.

2.8.2 VALIDATION & VERIFICATION STAGE LIFE CYCLE

Data Validation and Verification Stage inputs and deliverables:

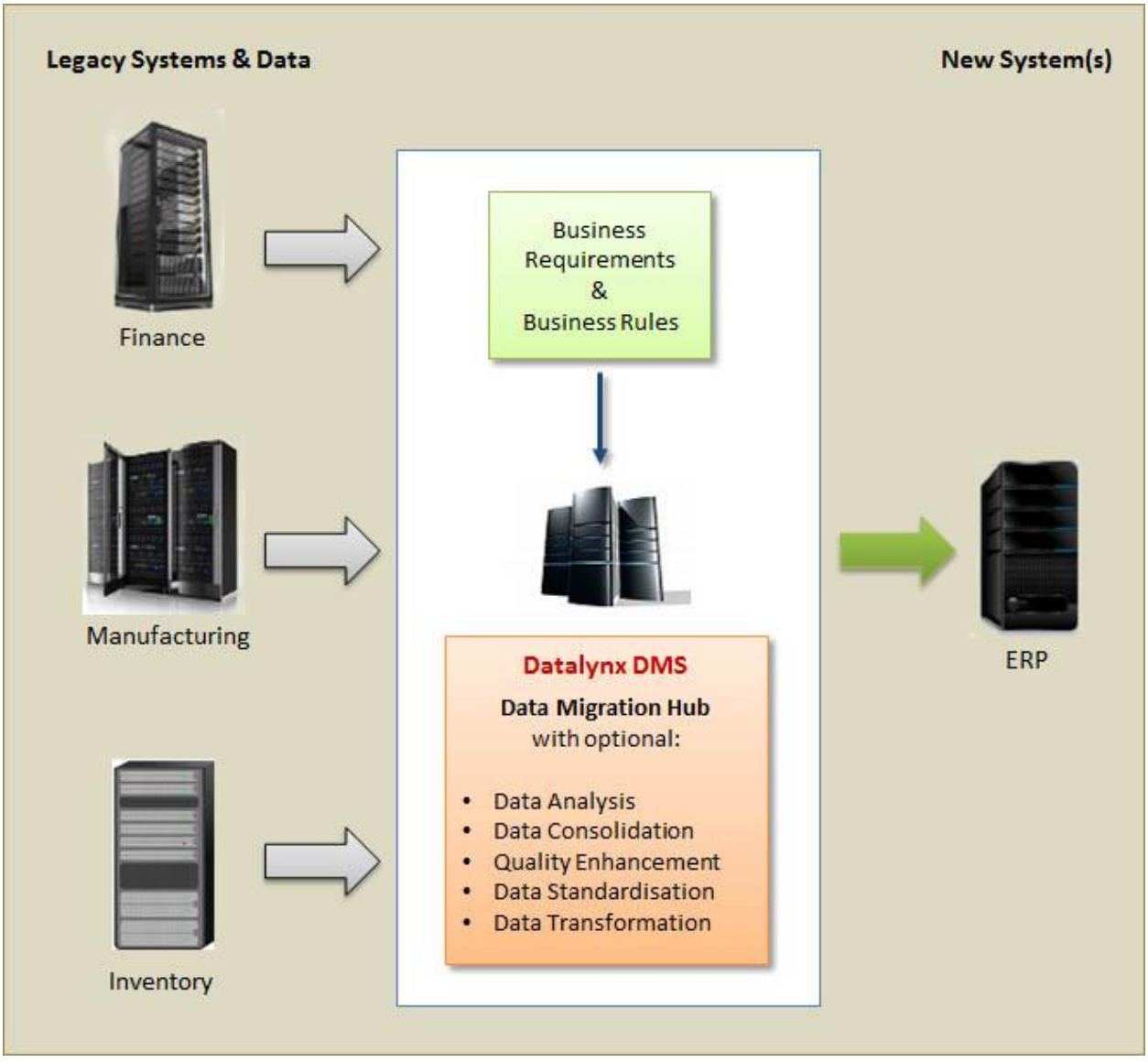
Input	Data Management Systems	Deliverables
• Source Database (staging) • Cleansed Database (Interim) • Transformed Database (Holding) • Migrated Database (System)	• Datalynx Data Xchange • Datalynx Data Xplorer	• Data profiling reports • Data cleansing reports • Summary report • Database metadata and record counts at the end of each stage of the process.

© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

2.9 DATALYNX DATA MANAGEMENT SUITE

The Datalynx DMS is the ideal software for configuring a powerful and flexible Data Migration Hub.

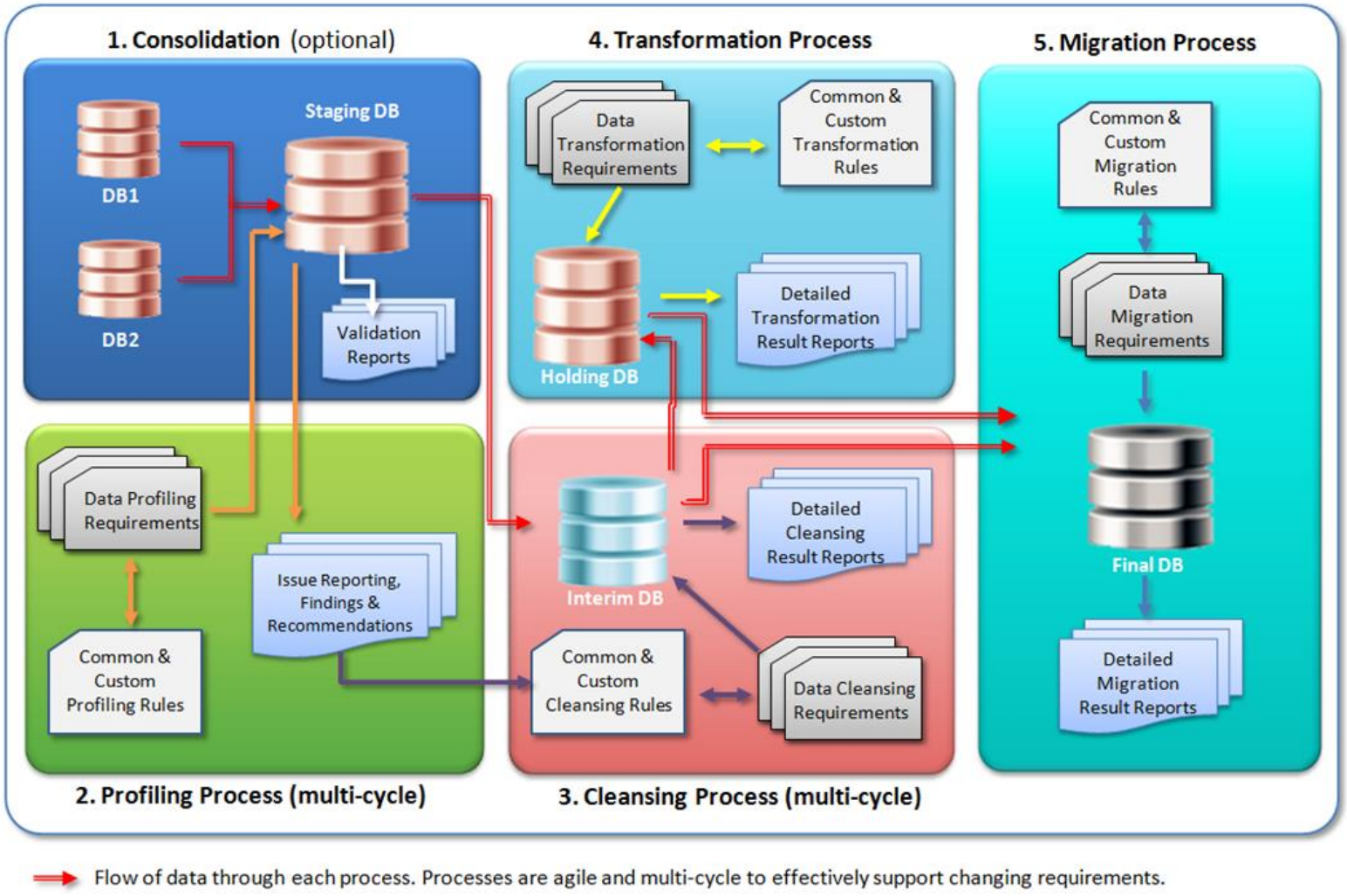
Offering inbuilt connectivity to a broad range of databases and structured data sources, enhanced ETL / ELT, a sophisticated rules-engine and integrated auditing / verification reporting, the DMS supports all your data migration requirements. In addition to data extraction, data analysis, transformation and migration the Hub's optional capabilities include data consolidation, data cleansing and data archiving.



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au

3 APPENDIX A – PCTM METHODOLOGY

The following diagram illustrates Datalynx's PCTM methodology for cleansing & migrating data:



© Copyright 2014 - Datalynx Pty Ltd. All rights reserved. www.datalynx.com.au