

State of California



FI\$Cal Preliminary Data Cleansing Guidelines

Version 2.0

March 28, 2013

Revision History

Date	Version	Last Updated By	Status/Comments	Change Approved By
07/16/2012	12011v1	David Porter	Initial version	N/A
07/20/2012	12011v2	David Porter	Updated the document to describe the general data cleansing approach.	N/A
11/28/2012	12011v2	Matt Gorkiewicz	Updated the document based on review.	N/A
11/29/2012	12011v2	Matt Gorkiewicz	Updated the document based on comments from Tej Kohli.	N/A
12/3/2012	12011v2	Matt Gorkiewicz	Updated the document based upon review with tech team.	N/A
12/5/2012	12011v2	Matt Gorkiewicz	Updated the document based upon review with Dave Porter.	N/A
12/11/2012	12011v2	Matt Gorkiewicz	Updated the document based upon review with Kelly Holtz.	N/A
12/18/2012	12011v2	Matt Gorkiewicz	Updated the document based upon review with deputy directors.	N/A
1/24/2013	12011v2	Matt Gorkiewicz	Updated data cleansing elements after resolution of ODMF #79 for conversion scope	N/A
1/28/2013	12011v2	Matt Gorkiewicz	Updated the document based upon review for scope.	N/A
2/2/2013	12011v3	Dave Porter	Updated the document based on review	N/A
2/6/2013	12011v4	Harnaik Kang	Reordered Sections based on review with Dave Porter.	N/A
3/4/2013	12011v5	Susan Dyckes	Updated document text.	N/A
3/11/2013	12011v7	CMO updates	Reordered sections, added content regarding MDW and Conversion Scope	N/A
3/27/13	12011v7	Ruth Ella George	Review	N/A
3/27/13	12011v7	Comms Unit	Edits	N/A
3/28/13	12011v8	CMO	Edits incorporated Version 2.0	N/A

This page intentionally left blank.

Table of Contents

REVISION HISTORY	2
1.0 PURPOSE OF THIS DOCUMENT	5
2.0 DATA CLEANSING OVERVIEW	6
3.0 PROCESS FOR PRELIMINARY DATA CLEANSING.....	7
4.0 DATA CLEANSING OBJECTS	9
5.0 DATA CLEANSING GUIDING PRINCIPLES & RESPONSIBILITIES.....	11
6.0 GENERAL DATA CLEANSING ISSUES.....	12
7.0 TIMELINE.....	14

1.0 Purpose of This Document

The purpose of this document is to provide departments with an approach for cleansing data residing in Legacy Systems in preparation for conversion to FI\$Cal. In addition to describing the most common data issues that are seen in Legacy Systems, an overview of the type of actions a department can perform to initially cleanse its data is provided. This approach assumes that preliminary data cleansing will be performed by departmental staff, and will be performed in the Legacy Systems. The work to perform the data cleansing will need to be a collaborative effort between both business and technical staff at the departments.

This document describes general steps and guidelines for preliminary data cleansing. The exact method for cleansing a department's data will vary depending on the data type, volume and storage method.

2.0 Data Cleansing Overview

To transfer Legacy data, it must be converted into a format that is useable by FI\$Cal. Data cleansing is the process of reviewing, updating and maintaining Legacy System data so that it can be converted into the FI\$Cal solution without intervention at final conversion time. Data cleansing is one of the most important processes needed to prepare for data conversion.

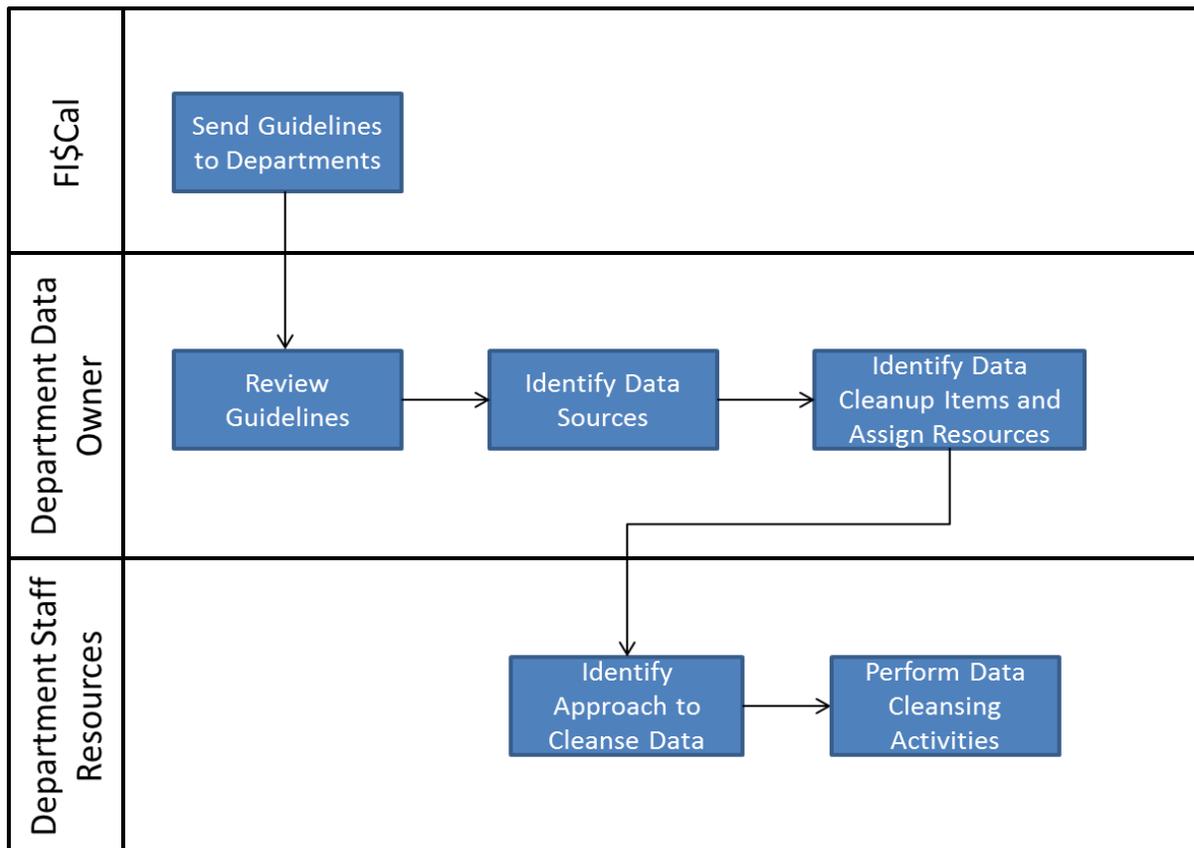
Tasks within the Master Department Workplan (MDW) address data cleansing and data conversion activities. This Data Cleansing Guideline document is a tool to assist departments in developing a plan to complete the data cleansing tasks in the MDW. Departments are responsible for cleansing their own data and will need to allocate resources to perform the data cleansing. Ideally, a department's business and technical Subject Matter Experts (SMEs), who are most familiar with the Legacy application, will either perform or oversee the data cleansing process.

The preliminary cleansing efforts should address the data cleansing issues described in Section 6.0, General Data Cleansing Issues, and should target data required by FI\$Cal. Section 4.0 provides a summary of the Data Cleansing Objects. The data cleansing efforts should focus on the most recent data first and, then proceed to older data.

The "Conversion Scope for FI\$Cal Project" (Scope) document has been provided with the FI\$Cal Preliminary Data Cleansing Guidelines. The Scope contains a high-level summary of the scope of the data that will be required in FI\$Cal. This document provides a preliminary inventory of Legacy System Data Objects for all waves in FI\$Cal, and lists data objects by category, along with the source, and years of data to be converted. As FI\$Cal progresses through the design and build process, each Legacy System Data Object required from departments for conversion will be addressed via tasks in the MDW. Note that not all data objects in the Scope may apply to every department.

3.0 Process for Preliminary Data Cleansing

The general data cleansing process is diagrammed below and described on the following page:



The general data cleansing process is as follows:

- 1) Departments receive and review the Preliminary Data Cleansing Guidelines from FI\$Cal.
- 2) Departments identify the primary source and data owner for each of the areas in Section 4.0 Data Cleansing Objects.
- 3) For each Legacy System identified, department resources walk through the list in Section 6.0 General Data Cleansing Issues and determine which situations apply to that system. Department resources should also identify any known data cleansing items that are specific to that Legacy System.
- 4) For each issue in the Legacy System, the data owner determines the appropriate departmental resources to perform the data cleansing.
- 5) The assigned department resource determines the best approach and plan for correcting the issues based on the volume and complexity of the data cleansing effort, as well as the tools native to the Legacy System which are available. The cleansing should occur starting with the most current data and, then proceed to older data.
- 6) As necessary, consult with your department's business and technical SMEs, your Readiness Coordinator, or email FI\$Cal at fiscal.cmo@fiscal.ca.gov for guidance on questions related to the preliminary data cleansing.

4.0 Data Cleansing Objects

Legacy Data Objects need to be identified, reviewed and cleansed by departments based upon the preliminary data cleansing process described in Section 3.0 of this document.

Further details will be provided during the design and build process for each data object needed for FI\$Cal, as the conversion layouts become available (refer to Section 7.0 Timeline for details about the timeframes of the phases). Not all data objects will apply to all departments.

To begin data cleansing activities, departments should identify the primary source and data owner for each of the areas below. Additional Legacy System Data Objects can be found in the Conversion Scope for FI\$Cal Project document.

OBJECT	CLEANSING CONSIDERATIONS
Vendor Data	<ul style="list-style-type: none"> • Active vendors that are currently being used for Accounts Payable processing • Ensure name, address, Tax Identification Number (TIN) / Social Security Number (SSN), vendor type are populated
Outstanding Requisitions	<ul style="list-style-type: none"> • Requisition process should be completed in Legacy Systems prior to cutover (to FI\$Cal) to create the purchase order in the Legacy System • Close out any requisitions that are no longer active
Outstanding Purchase Orders / Contract Encumbrances / Outstanding Encumbrance Documents	<ul style="list-style-type: none"> • Include only open purchase orders (encumbrances) • Close historical purchase orders, encumbrances, and contracts that are no longer being used • Ensure vendor, purchase order amount, and accounting information are populated and accurate
Outstanding Vouchers / Claim Schedules	<ul style="list-style-type: none"> • Pay or close outstanding historical vouchers which are still open • Includes Office Revolving Fund (ORF) open advance transactions (Salary, Travel, Miscellaneous)
Open Accounts Receivable	<ul style="list-style-type: none"> • Only open receivables will be converted • Ensure customer, amounts, dates are populated

OBJECT	CLEANSING CONSIDERATIONS
Customers	<ul style="list-style-type: none"> • Include active customers with open receivables or common customers who the department will invoice regularly • Ensure customer name, and address are populated
Asset Information (refer to SAM 8600 Property Accounting)	<ul style="list-style-type: none"> • Assets include infrastructure, land, buildings, improvements, machinery, furniture, tools, and other intangibles • Dispose of any assets that are no longer in use • Ensure Asset ID tag, accounting information, asset location, asset classification are populated: for capital assets, and depreciation (begin date, asset life, amount)
Current Projects	<ul style="list-style-type: none"> • Projects which are still open and active • Projects that are both funded and not funded by bonds • Ensure the beginning balance of the project is available • Ensure inception to date transactions and dates are populated • Close out completed projects
Grants/Grantees	<ul style="list-style-type: none"> • Open and active grants • Ensure grant dates and grantee information are populated • Ensure transactions for the grant are available and populated
Financial Balances	<ul style="list-style-type: none"> • Accounting Data – Beginning balances, current and prior fiscal years (ledger level, not journals) • Real/Balance Sheet Accounts – Beginning balances, current and prior fiscal years (ledger level, not journals)
Budgets	<ul style="list-style-type: none"> • Appropriations/Allotments – Beginning balances, current and prior fiscal years (ledger level, not journals) • Current and prior Fiscal Year • Data primarily from DOF budget system

5.0 Data Cleansing Guiding Principles and Responsibilities

The Data Cleansing process will follow a series of guiding principles and responsibilities. Listed below are the primary principles and assumptions for the process:

- 1) Legacy data must undergo data cleansing to improve quality, minimize data integrity issues, reduce data volume, and decrease run time for conversion extract programs.
- 2) Departments will be responsible for cleansing Legacy System data before it is loaded into the staging area for FI\$Cal.
- 3) Departments will be responsible for supplying the resources to complete data cleansing activities in the Legacy Systems.
- 4) Departments are responsible for ensuring that extracted data has been verified before it gets loaded into the FI\$Cal environment, so that only the data intended for conversion is sent to FI\$Cal.
- 5) Departments are responsible for identifying a data owner for each Legacy application through the Legacy System Analysis (LSA) system validation. A data owner will be assigned for each conversion effort and he or she will be responsible for the completeness and accuracy of the source data to be converted.
- 6) FI\$Cal will provide the FI\$Cal data requirements and the corresponding support to help departments understand the FI\$Cal data fields for each stage of the process.

6.0 General Data Cleansing Issues

The following table depicts the common issues encountered within Legacy Systems that require cleansing prior to data conversion. The issues listed will not apply to all systems in all departments, but reflect the trends generally seen. The table defines the potential issue and an explanation of the issue. In addition to the issues highlighted in the table, there may be data cleansing issues unique to a departmental Legacy System. Also, there may be pre-existing data issues in the Legacy Systems that require data cleansing. The department will need to work to identify and resolve these issues.

ISSUE	EXPLANATION
Duplicates	<p>The same data entity (for example, asset, vendor and customer) is named two or more times in the same system.</p> <p>Example: The same vendor is listed multiple times in the vendor file.</p>
Obsolete or Inactive Records	<p>Data that is not up to date or no longer active. Obsolete data is not needed in FI\$Cal. Through the normal course of business, records should be inactivated / closed when appropriate.</p> <p>Example: Vendors who are out of business or similarly inactive. Requisitions / Purchase Orders which are closed.</p>
Inconsistent Data	<p>Inconsistencies that are related to typing or data entry errors. The same value is represented in a field in multiple methods. A common field is used differently across the system. The State field is populated with "CA" in some systems and "California" in others.</p> <p>Example: The values of "Y," "YES," and "Yes" are all used in the same field to represent the same value.</p>
Inconsistent Similar Data	<p>Similar data entered into separate or independent systems.</p> <p>Example: Departments defining projects in multiple systems. Same type of data (project related) is entered into different systems, but since it is not validated against each other or a central system, the data format is different.</p>

ISSUE	EXPLANATION
Incomplete Records	Missing data in current Legacy System. Required fields are missing from either a system perspective or fields that are required from a business perspective, but not enforced programmatically by the system.
Parent/Child Relationship Missing	A required parent/child relationship between data in the system does not exist. Example: Header record with no detail lines for a transaction.

7.0 Timeline

FI\$Cal will progress through a series of phases as the system is designed and built. The chart below illustrates the different phases of the conversion and highlights the primary conversion and data cleansing related tasks. Preliminary data cleansing efforts will occur in parallel with the Legacy Systems Analysis (LSA) outreach and initial conversion or interface recommendations.

ACTIVITY	DATA CLEANSING DETAILS	PHASE
LSA "Collect and Validate System Information"	Preliminary Data Cleansing Guidelines document provides common data cleansing situations encountered in systems and steps to resolve.	Analyze
Legacy Conversion and Interface Analysis	The functional design produced out of the Analyze phase will provide additional details about the areas and scope of the conversion, including additional data that will potentially need to be cleansed.	Design
Conversion Data Mapping and Build	Detailed layouts of the conversion files which will need to be mapped and the associated data cleansing instructions.	Build
Mock Conversion Testing	Full testing of the conversion process with Legacy data, providing additional data cleansing items based upon the results.	Test
Transition to Production	Final conversion into FI\$Cal production system, data cleansing as required during and following conversion.	Deploy

The chart below calendars the Wave 1 activities and phases:

Wave 1 Activities	2012			2013												2014						
	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	JUL	AUG	SEP	OCT	NOV	DEC	JAN	FEB	MAR	APR	MAY	JUN	
Phase	Analyze			Design					Build				Test						Dep			
Collect and Validate System Information																						
Legacy Conversion and Interface Analysis																						
Conversion Data Mapping and Build																						
Mock Conversion Testing																						
Transition to Production																						