ctcLink Data Cleansing Instructions

College Student Bio/Demo Data Cleansing

The purpose of this document is to provide colleges instructions on how to begin data cleanup activities in preparation for deployment of the PeopleSoft solution. Included are instructions on generating an error file for data clean-up and guidance on how to correct errant data to enable conversion to PeopleSoft.

2013

ctcLink Project Team

State Board for Community and Technical Colleges

9/10/2013

Table of Contents

[1.0 Understanding the Need for Data Clean-Up 3](#_Toc366667244)

[1.1 Executive Summary 3](#_Toc366667245)

[1.2 Overview of Data Clean-Up Process 3](#_Toc366667246)

[2.0 Generating the Error List 4](#_Toc366667247)

[2.1 Running DATAEXPRESS to Produce the Excel (CSV) Input File 4](#_Toc366667248)

[2.2 Using the Excel Macro-Enabled Template 4](#_Toc366667249)

[3.0 Data Clean-up Instructions 6](#_Toc366667250)

[3.1 Element: Student Name 6](#_Toc366667251)

[3.2 Element: SSN 6](#_Toc366667252)

[3.3 Element: Address 7](#_Toc366667253)

[3.4 Element: E-mail Address 8](#_Toc366667254)

[3.5 Element: Telephone Number 8](#_Toc366667255)

[3.6 Element: Handicap Status 9](#_Toc366667256)

[3.7 Element: Emergency Contact 9](#_Toc366667257)

[3.8 Element: Birth Date 10](#_Toc366667258)

[3.9 Element: Citizenship Status 10](#_Toc366667259)

# Understanding the Need for Data Clean-Up

## 1.1 Executive Summary

Colleges are being asked to review data error conditions in the legacy system and make decisions regarding corrections. As the project progresses through the various conversion programs a clean-up guide will be produced and distributed by data area. Data owners/sponsors will be responsible for ensuring that data is cleansed (at the source) in coordination with the ctcLink Project Team. College executives are asked to support the time and effort expended by college Subject Matter Experts (SMEs) necessary to complete the following:

* Review the error conditions and make decisions regarding how to handle the error
* Communicate those decisions to the ctcLink Project Team
* Correct erroneous legacy data conditions (As needed – priority order by critical data elements)
* Communicate with the ctcLink Project Team when clean-up is complete for a particular data element.
* Recognize business practices that may be contributing to data cleanliness and either change them or commit to continued data cleansing to support the data conversion testing process.

The goal of data clean-up activities at the college is to ensure data can be loaded into the PeopleSoft system both during data conversion testing and at deployment or “GO LIVE.” For BIO/Demo data an additional step POST “GO LIVE” is recommended for students to validate their information and update their personal information in PeopleSoft Student Self-Service.

Pillar specific questions regarding data clean-up activities can be directed to the ctcLink Project Team pillar leads. General questions or concerns about the data clean-up process can be addressed to Tara Keen, Functional Application Manager or Ray Gartner, Technology & Integration Manager.

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| **Campus Solutions** | **HCM** | **Finance** | **DataExpress** | **Functional Mgr.** | **Technical Mgr.** |
| Betsy Abts | Rita Lauzon | Cindy Schaffer | SBCTC-ITD Support | Tara Keen | Ray Gartner |
| babts@sbctc.edu | rlauzon@sbctc.edu  | cschaffer@sbctc.edu | support@sbctc.edu | tkeen@sbctc.edu  | rgartner@sbctc.edu |
| 425.803.5359 | 425.803.9723 | 425.803.5358 | 425.803.9721 | 425.803.9736 | 425.803.9713 |

## 1.2 Overview of Data Clean-Up Process

One of the most time-consuming elements of any data conversion project is the effort necessary to ensure source data (legacy system data) is clean and in a format that can be converted into the target system (PeopleSoft). This effort is an iterative process of executing a test conversion, finding all data that generates an error, determining what is needed to resolve the error and running the conversion again to validate the resolution. Some data adjustments can be made within the conversion program, but the best approach is to clean the data at the source (existing system).

This document provides instructions for running DataExpress procedures to show a list of records that contain errors or invalid data. The end result of this process produces an Excel spreadsheet with filter capability and sorted by the Comment column which describes the type of error.

# 2.0 Generating the Error List

## 2.1 Running DATAEXPRESS to Produce the Excel (CSV) Input File

The instructions below describe how to run a DataExpress procedure to generate an exception file and download the results for use by the Excel template. A recommendation is provided in step 6 of the instructions for a possible naming convention.

1. Log on to your HPUX account and access DataExpress.
2. From the DataExpress Main Menu, choose option 2:

3. Type the name of the DataExpress procedure and press return. See example below.

4. Specify the catalog location: 
5. After the procedure has executed, the following prompt will be displayed. Press return.
 
6. When prompted for a file name, you may specify a file name and location of your choosing. The filename shown in the example below “QS1001X” was chosen to represent that file is output from the QS1001R ‘report’ in an ‘Excel compatible’ format, represented by the “X.” If running this report multiple times it is recommended to adopt a QS1001X\_<*date*> format.


## 2.2 Using the Excel Macro-Enabled Template

Using the example file naming convention described in section 2.1, step 6, the instructions below describe how to open the Excel template associated with the exception file. The macro in the template will import the DataExpress downloaded file and format it for college use.

1. Open your internet browser and go to: <https://ctclink.sharepoint.com/FunctionalProcesses/Student/SitePages/Home.aspx>
2. Find the SME gathering agenda dated 8/22/13 and select the meeting title and find the link to the attachment.
3. Click on the link for the Exception\_Template.xltm Excel Macro-Enabled Template. When the template opens, it will be blank.
4. With the blank template open, execute the macro by holding down the Ctrl button and pressing the lowercase letter “i” (Ctrl+i).

5. Your Windows Explorer window will open (example below). Navigate to the location where you saved the DataExpress output file and double-click on the file name (For example, **QS1001X**).



6. After double clicking on the filename in step 5, the macro will continue by opening the file and formatting the results into a sorted and filtered output. Colleges may wish to save the results locally for future dissemination and reference. Keep in mind, if running this report multiple times it is recommended to adopt a QS1001X\_*<date*> format.

# 3.0 Data Clean-up Instructions

The following instructions are listed by the priority of importance. Some data elements will prevent the loading of a record during data conversion. Those clean-up activities will have a higher priority. Other data elements will have priority due to the impact on the ability for Search/Match to avoid duplicating records.

**Important Note: Each time an element cannot be loaded it reduces the options for the search/match process.**

Once you have opened the excel file via the template, please start by filtering on the elements in the order they are listed below.

## 3.1 Element: Student Name

Issue: In order to load a Student record into PeopleSoft, it must contain both a first and last name. This is derived from the legacy STU-NAME field. This is done by unstringing STU-NAME delimited by a single space. Therefore STU-NAME must meet the following criteria:

* Must contain a minimum of two words separated by a single space.

Error Message:

MISSING FIRST NAME
MISSING LAST NAME
MISSING NAME

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Ensure there is no more than 1 space between last and first name
* Ensure there are no leading spaces in the student name field
* Note any students whose last name occupies the entire length of the student name field (has no first name in student name)-These names will be manually populated into PeopleSoft by ctcLink

## 3.2 Element: SSN

Issue: In order for SSN to be loaded and used as part of the student’s bio/demo record, the SSN must meet the following criteria:

* Must be 9 digits
* Character 1 cannot be “9”
* Characters 1-3 cannot be “000” or “666”
* Characters 4-5 cannot be “00”
* Characters 6-9 cannot be “0000”

Error Message:

 INVALID SSN

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Remove invalid SSN’s
* Keep invalid SSN’s in legacy with the knowledge that the SSN will not load to PeopleSoft, however, the student name and all other valid values will load.
* If the SSN’s were entered as a fake on purpose, explain why and trouble-shoot your options with functional analysis.

## 3.3 Element: Address

Issue: In order for Addresses to be loaded and used as part of the student’s bio/demo record, the address must meet the following criteria:

* Street cannot be blank
* Must not contain the ‘@’ character
* City cannot be blank
* State must be a valid 2 character USA state code
* Must be a valid state for the zip code on the Zip Code Table (screen ZX0020)
* Zip Codemust be a valid zip code on the Zip Code Table (screen ZX0020)

Error Message:

INVALID STREET
MISSING STREET
MISSING CITY
INVALID STATE
INVALID STATE FOR THE ZIP CODE
MISSING STATE
INVALID ZIP
MISSING ZIP

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* If possible add missing value
* It may require that the zip code table be corrected if zip can be validated with USPS zip look-up.
* If may require that the zip code table be corrected with the valid state code (should also be validated with USPS look-up)
* Keep invalid Address in legacy with the knowledge that the address will not load to PeopleSoft, however, the student name and all other valid values will load.

## 3.4 Element: E-mail Address

Issue: In order for Email Address to load, it must meet the following criteria:

* Must contain a ‘.’ (dot) , ‘@’ sign, and ‘.’ after the ‘@’ sign
* Must not contain an ‘@’ at the end of the email address

Error Message:

 INVALID EMAIL

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* If possible add missing value
* Keep invalid Email Address in legacy with the knowledge that the email address will not load to PeopleSoft, however, the student name and all other valid values will load.

## 3.5 Element: Telephone Number

 Issue: In order for Phone Numberto be loaded and used as part of the student’s bio/demo record, the telephone number must meet the following criteria:

* Must contain 7 or 10 digits.

Error Message:

 INVALID PHONE NUMBER

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* If possible add missing value
* Keep invalid Telephone Number in legacy with the knowledge that the telephone number will not load to PeopleSoft, however, the student name and all other valid values will load.

## 3.6 Element: Handicap Status

Issue: In order for Handicap Statusto be loaded and used as part of the student’s bio/demo record, the handicap status must meet the following criteria:

* Must be either Y or N

Error Message:

 INVALID HANDICAP STATUS

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* Remove invalid value
* Keep invalid Handicap Status in legacy with the knowledge that the handicap status will not load to PeopleSoft, however, the student name and all other valid values will load.

## 3.7 Element: Emergency Contact

Issue: In order for Emergency Contact infoto be loaded and used as part of the student’s bio/demo record, the emergency contact info must meet the following criteria:

* Must contain both a first and last name a minimum of two words separated by a single space.

Error Message:

MISSING LAST NAME
MISSING FIRST NAME

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and hse the following advice for data cleansing.

* Ensure there are no more than 1 space between last and first name
* Ensure there are no leading spaces in the student name field
* Replace invalid value with a valid value
* Remove invalid value
* Keep invalid Emergency Contact in legacy with the knowledge that the emergency contact will not load to PeopleSoft, however, the student name and all other valid values will load

## 3.8 Element: Birth Date

Issue: In order for Birth Dateto be loaded and used as part of the student’s bio/demo record, the birth date must meet the following criteria:

* Birthdate be a valid date
* Birthdate cannot be a date in the future

Error Message:

 INVALID BIRTHDATE

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* Remove invalid value
* Keep invalid birth date in legacy with the knowledge that the birth date will not load to PeopleSoft, however, the student name and all other valid values will load
* Known problem of valid birth dates that are less than 12/31/1919 hitting the error file can be ignored

## 3.9 Element: Citizenship Status

Issue: In order for Citizenship Statusto be loaded and used as part of the student’s bio/demo record, the citizenship stat must meet the following criteria:

* Must be a va*lid Visa Type OR the following*
* Fir*st character must be alphabetic*
* First character must not be N S U V W X Y Z; except ‘ Y ‘ ‘N ‘, ‘Z ’ or ‘ZZ’ is ok

Error Message:

 INVALID CITIZENSHIP STATUS

Error Handling

A ctcLink Functional Analyst has reviewed the error records for your college and has the following advice for data cleansing.

* Replace invalid value with a valid value
* Remove invalid value
* Contact SBCTC about adding the value to Table 3