eScholar Data Manager™

User Guide

Version 2021

Last Updated: September 20, 2021

CONTENTS

ABOUT THIS MANUAL

This manual serves as a User Guide for the eScholar Data Manager™ application. The eScholar Data Manager™ application is the web based interface to the eScholar Data Complete Data Warehouse®.

This manual includes some notes that are designed to highlight important information. The following Notes indicators are used:

> **NOTE:** *This box mention things that require special attention. The symbol to the left indicates an important note to remember.*

DATA FLOW

The eScholar Data Manager™ application is the web based interface to the eScholar Data Complete Data Warehouse®. The data flow into the warehouse is represented in the diagram below:
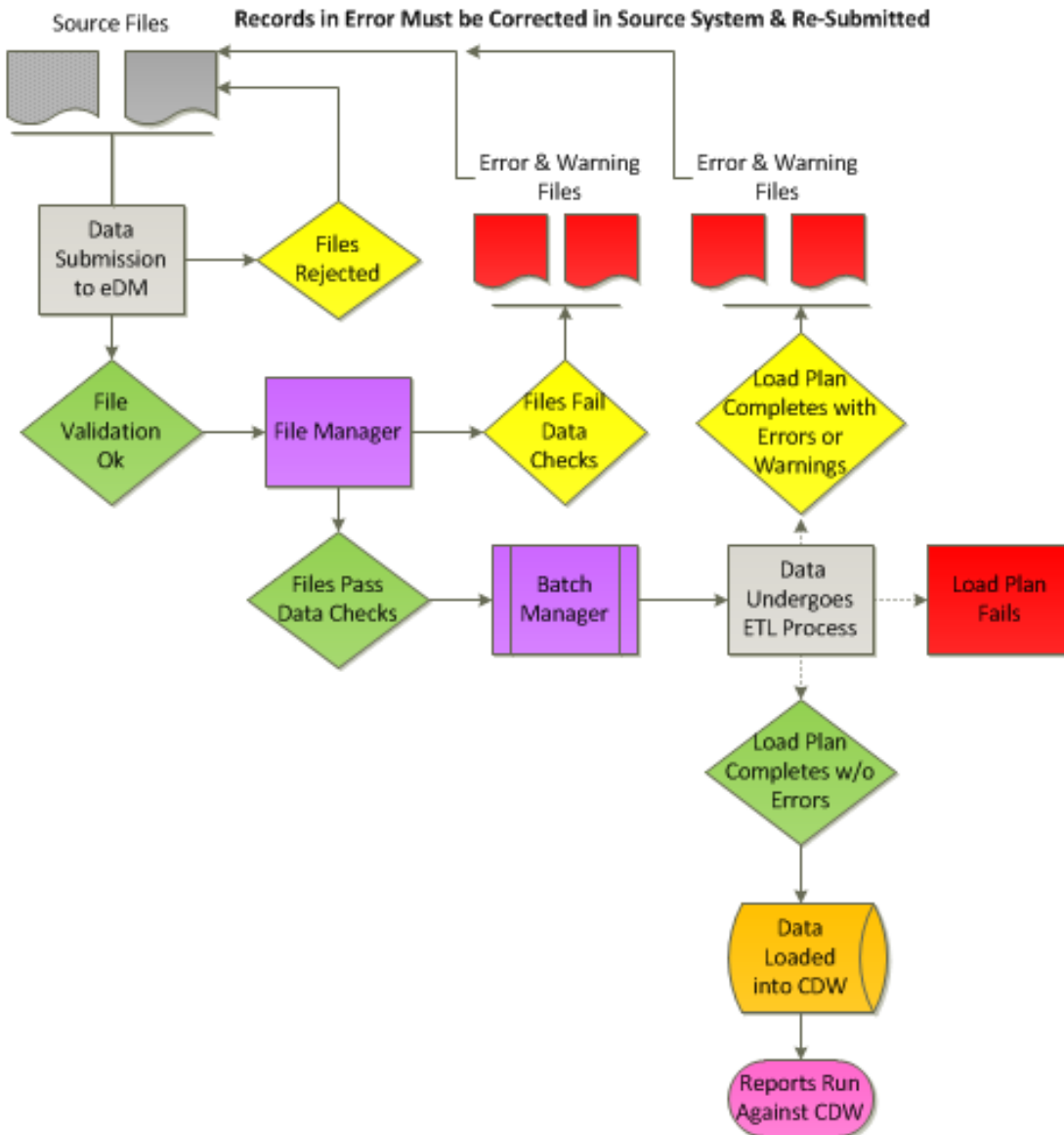
DATA QUALITY GATES

Within eScholar Data Manager™ application there are three data quality gates—File Manager, Data Quality Engine and ETL Process.

Gate #1 – File Manager

In the File Manager files are validated for:

1. Data Types – validates dates are valid dates, numerics, and strings.
2. Data Length – validates the field length does not exceed the maximum allowed characters.
3. Required Validation – mark the template field is required. If a record is uploaded without a value that record fails validation.
4. No Value Allowed – mark the template field that no value is allowed. If a value is supplied the record fails validation.
5. Range – for any date or number the administrator can specify a range that is valid for that template field.
6. Valid Value – for the template field a list of values that are allowed for that field. If a value is supplied and does not match a value on the list, the record fails validation.
7. Format Mask – for the template field if a value is supplied it must match the expression. This is similar to a regular expression.
8. Snapshot Date Validation – for the template field if a value is supplied it must match the snapshot date of the data collections.
9. Student ID – for the template field if a value is supplied it must exist in the table of valid student IDs. This table is typically populated from eScholar Uniq-ID® application.

> *NOTE: In File Manager, any record with one or more data quality errors are removed from file. When the validation process is completed, there will be a "Clean File" - all records in the clean file has zero data quality checks. And, there will be an error file containing all the errors found in the source file.*

Gate #2 – Data Quality Engine

Once files are batched, they are passed to the batch process. The batch process will check if there are any Advanced Rules associated with the data collection. If there are one or more rules, the incoming batch data is validated against the advanced rules. Some examples of Advanced Rules are:

- Any student with grade level equal to "PREK" the student age is less than or equal to four years old.
- Any student with grade level equal to "PREK" then the student must also be enrolled in at least one PREK program.
- A year over year check, any student reported as graduated the prior year is not enrolled in the current year.

If the data passes the Advanced Rules validations, the batch data is moved on to the ETL process.

Gate #3 – ETL Process

In the batch process, once data has passed the first two gates, the data moves on to the last step the ETL process.
In the ETL, files are validated for:
1. Referential Integrity – validates that for any record any pre-requisite record already exists. For example, if you provide an assessment record for a Student ID not in the Student table this will be logged as an error.
2. Valid Values – valid values validations can also be applied in the ETL process as well they can be setup in the File Manager.
3. Duplicate Checking – the ETL process checks for duplicates based on the logical key of the template row.

UPLOADING FILES INTO ESCHOLAR DATA MANAGER

Preliminary Requirements

There are two main requirements for data files that are to be loaded into the data warehouse via eDM:
- The files must be formatted as eScholar templates.
- The files must adhere to a naming convention.

eScholar Templates

All data files to be loaded into the data warehouse must be in a format that the data warehouse ETL process can understand. The format of the files depends upon the type of data being loaded, but these formats are well defined in the eScholar templates. The system accepts several different types of files (comma or tab delimited) so long as they conform to the layouts specified by the eScholar templates.

For more information about the various file formats accepted by eDM and the eScholar Complete Data Warehouse®, please consult the eScholar template documentation on the eScholar support site.

eDM File Naming Convention

In order to process data files, eDM needs to know a few things about the data contained in the files that are submitted to it. Namely, eDM needs the following information:
- The agency/source to which the data belongs
- The type of data contained in the file
- The version of the data
- The format of the data within the file

eDM determines all of these things based upon the name of the data file. Specifically, eDM expects all files loaded to conform to the following naming convention:

<district code >_<template type>_<timestamp>.<file type>

Where
- District code is the assigned identifier for the district whose data is contained in the file

- Template type is the type of data being loaded (e.g. STUDENT, SCHOOL_ENROLL, etc.)
- Timestamp is a date in YYYYMMDDHHMM format (e.g. 201006021015)
- File type is the layout of the file (CSV for comma-separated, TAB for tab files)

For example, this would be the name of a data file containing student data for district 123 in comma-separated format:
2251001_STUDENT_201111211104.CSV

Uploading Data Files

Uploading data files is one of the primary purposes of eDM. There are two ways to upload files into eDM: via the web browser and via automation.

Uploading via a Web Browser

The usual method of uploading files into eDM is via a web browser from the File Manager screen. Simply click on the Upload link on the left menu, and you will be presented with a screen named File Upload that allows you to select files on your local computer for submission to eDM. To upload any file, you must:
1. Select the data collection the data file(s) are intended for.
2. Select the file(s) to upload.
3. If your file(s) contain a header row click the checkbox to indicate the file contains a header.
4. Click the upload button.

✍ | **NOTE:** *Please be patient. Once your file(s) have been uploaded, you will be redirected to the File Manager page.*

Uploading via Automation

There are two automation options for loading using eScholar Data Manager™.
1. FTP/Folder Automation – a file location can be setup on the network. Template files can be dropped in the designated folder, then loaded on a schedule.
2. REST base API – eScholar has a REST based API for data loading. To use the REST base API please contact your account manager.

FILE MANAGER

File Manager Page

Data files can be broken down into 2 categories within eDM: those that have been uploaded to the system but not yet put into a batch (and therefore not yet loaded into the data warehouse), and those that have been put into a batch (batching files is covered later in this document). This section covers the File Manager, which can be viewed as a holding place for files that have been uploaded to the system, but have not yet been loaded into the data warehouse.

Sort Order

The files displayed on the File Manager screen are sorted by their associated ID (which is assigned when the file is uploaded) in reverse order. As a result, the most recently uploaded files appear at the beginning of the list, while the oldest files appear at the end. Files can also be sorted by File Name, Upload Time, and Status by clicking on the column headers.

File Validation Statuses

eDM™, and the Complete Data Warehouse®, perform numerous validations on uploaded data before it is actually placed into the data warehouse. These validations come in two stages. The first stage occurs when a file is uploaded to eDM. This is called pre-ETL validation. The second stage of validation occurs when a file is run through the ETL process, during which numerous validations are performed on the data.

As a result, after a file is uploaded to eDM, it will have validation rules applied to it which will affect the display on the File Manager screen.
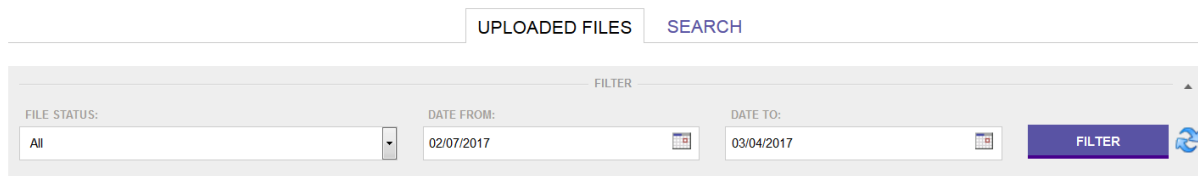
🔧	**Received Status** - means that eDM has received a file but has not completed pre-ETL validation on the file.
📄❌	**Failed Validation Status** - The data file has completed validation, but there were enough errors found to cause the file to fail validation. Files that fail pre-ETL validation cannot be placed into a batch.
📄⚠️	**Validation Warning Status** - The data file has completed validation, and some errors were found. However, the number of errors found was not enough to cause the file to fail validation. The results of validation for these files should be reviewed. Files that have validation warnings may be placed into a batch.
📄✔️	**Validation OK Status** - The data file has completed pre-ETL validation, and no errors were found. Files that have no pre-ETL validation errors can be placed into a batch.

✍️
NOTE: It is important to note that just because a data file completes pre-ETL validation with no errors does not mean that the data contained within the file has been loaded into the data warehouse. These files, like those that end up in Validation Warning status, must be placed into a batch before their data is loaded into the data warehouse.

Filtering Files

The list of data files displayed on the File Manager screen can be filtered by either the date they were uploaded to the system or validation status. Simply choose a desired date range and/or status from the options at the top of the screen and click on the Filter button.

UPLOADED FILES	SEARCH

	FILTER			
FILE STATUS:	**DATE FROM:**	**DATE TO:**		
All	02/07/2017	03/04/2017	FILTER	🔄

Searching for Files

There are four ways you can search for files in the File Manager:
- File ID – Search by Data File ID
- File Name – Search by Data File Name
- Status – Search by Data File Status
- Template Type – Search by Template Type like STUDENT, PROGRAMS FACT, etc.

SEARCH BY: ◉ FILE ID ◎ FILE NAME ◎ STATUS ◎ TEMPLATE

File ID :   913

☐ Search All Available Districts

SEARCH

MANUAL BATCHING FILES VIA FILE MANAGER

Overview

Batches are the mechanism that eDM uses to load your data into the data warehouse. A batch is simply a group of one or more data files that are processed as a unit through the ETL (extract, transform, and load) process, resulting in data being inserted or updated in your data warehouse.

Selecting Files to Include in the Batch

To create a batch, simply navigate to the **File Manager** page from the link on the left side of the screen. Next, select one or more files to be placed into the batch by clicking on the checkbox to the left of the file(s), then click on the **Add to Batch** button at the bottom of the screen. If you have more than one page of data files, you can navigate through the pages using the **First**, **Prev**, **Next**, and **Last** buttons at the bottom right of the page, select the file(s) that you want to add to the batch on each screen, and again click on the **Add to Batch** button on each screen.

Adding Comments to a Batch

You can add comments to any batch prior to processing it through the ETL process. Comments are optional, but can be very useful when trying to locate a batch that has been processed. Comments are limited to 255 characters.

Remove Files from Batch

If you decide that one or more data files in your batch should not be included in the batch, you can remove the file(s) from the batch by selecting the checkbox to the left of the file(s) and clicking on the **Remove From Batch** button toward the bottom of the screen.

Processing a Batch

Once you have selected the group of data files to be included in a batch and have (optionally) added an appropriate comment for the batch, the next step is to submit the batch for processing. To run the batch of data files through the ETL process, click on the **Process Batch** button on the View Batch screen. The Batch Manager screen will be displayed (see below).

NOTE: *You can create as many batches as needed. All batches are processed in the order that they are created, but only run one batch at a time.*

BATCH CART

Add comments associated with this batch (optional - limited to 255 characters):

Load Student and Programs

FILES IN BATCH

	FILE ID	TEMPLATE	COLLECTION	FILE NAME	UPLOADED TIME	STATUS**	
☐	9238	STUDENT	2017 Data Collection API	2251001_STUDENT_201111211104.CSV	2017-10-21 12:53	Validation OK	☑
☐	9239	SCHOOL ENROLLMENT	2017 Data Collection API	2251001_SCHOOL_ENROLL_201111211104.CSV	2017-10-21 12:53	Validation OK	☑
☐	9240	PROGRAMS FACT	2017 Data Collection API	2251001_PROGRAMS_FACT_201111211104.CSV	2017-10-21 12:53	Validation OK	☑

| DONE | REMOVE FROM BATCH | PROCESS BATCH |

FILE MANAGER – VIEWING ERRORS

Overview

There are two error views in File Manager. To View Errors:
1. Click the View Errors Button
2. The View Errors page displays the error counts and two options to view errors
   a. Error List – The errors are displayed in a detail list
   b. Highlighted Errors – HTML Grid of the template with highlighted errors

File Details Page

FILE DETAILS - FILE ID: 9241; FILE NAME: 2251001_STUDENT_201111211104.CSV

GENERAL INFORMATION | **VALIDATION INFORMATION**

VALIDATION START TIME:	2017-10-21 14:01
VALIDATION END TIME:	2017-10-21 14:01
TOTAL VALIDATION TIME:	0.803 seconds
NUMBER OF APPLIED VALIDATIONS:	256
VALIDATION STATUS:	Failed Validation

FILE CONTENTS

File Type	File Name	Records	Last Modified Date	Actions
Source File	2251001_STUDENT_201111211104.CSV	41	2017-10-21 14:01	VIEW SOURCE FILE

ERROR FILE CONTENTS

View By	Description	Records in Error	Total Errors Count	Actions	
Error List	This view displays the errors in a detail list	3	3	VIEW ERRORS	DOWNLOAD FILE (ZIPPED)
Highlighted Errors (Classic View)	This view displays the file contents in template format with highlighted errors	3	3	VIEW ERRORS	DOWNLOAD FILE (ZIPPED)

DELETE | DONE

Error List

FILE INFORMATION

FILE TYPE:	File Manager Error File
FILE NAME:	2251001_STUDENT_201111211104_FMGR_ERRORS.TAB

DOWNLOAD FILE (ZIPPED) | DOWNLOAD FILE | DONE

FILE CONTENTS

File Preview

Note: For performance reasons, only the first 100 rows of a file is displayed. To view the entire contents of larger files, you must download the file.

SOURCE RECORD NUMBER	FIELD NUMBER	FIELD NAME	FIELD VALUE	ERROR MESSAGE	ERROR MESSAGE VALUES
1	2	LOCATION CODE		Field Value Required	{"Location Code":""}
2	2	LOCATION CODE		Field Value Required	{"Location Code":""}
3	3	SCHOOL YEAR DATE	201Z-06-30	Field Value Not in ISO Format	{"School Year Date":"201Z-06-30"}

Highlighted Errors

FILE TYPE:	Validation Error File
FILENAME:	error.zip
TOTAL NUMBER OF RECORDS:	3

DOWNLOAD FILE (ZIPPED)   DONE

FILE CONTENTS

File Preview

Note: For performance reasons, only the first 100 rows of a file is displayed. To view the entire contents of larger files, you must download the file.

TIP - If a record appears in the error file below but no field is highlighted in red, it is likely that the record failed validation due to a template field that is not accounted for in the delimted source file. Please check the source file to ensure that all template field positions are accounted for in the record.

SOURCE RECORD NUMBER	#1 [DISTRICT CODE]	#2 [LOCATION CODE]	#3 [SCHOOL YEAR DATE]	#4 [STUDENT ID]	#5 [SOCIAL SECURITY NUMBER]	#6 [FAMILY NUMBER]	#7 [LAST NAME SHORT]	#8 [FIRST NAME SHORT]	#9 [MIDDLE INITIAL]	#10 [CURRENT GRADE LEVEL]	#11 [OBSOLETE]	#12 [OBSOLETE]	#13 [HOMEROOM]	#14 [BIRTH DATE]	#15 [GENDER CODE]
1	2251001		2017-06-30	1123899436			Zerphey	Alanna		01				2016-02-04	F
2	2251001		2017-06-30	1127455265			Monopoli	Amada		01				2016-02-07	F
3	2251001	1001	2017-06-30	1128563819			Gragson	Lakia		01				2006-01-27	F

DATA QUALITY ENGINE

Overview

The Data Quality Engine is embedded into the data flow process. If turned on, the DQE will execute rules after files are batched from the File Manager. If the batch fails any data quality rules, the rules and errors are displayed in the Batch Manager.

Viewing Data Quality Errors

If the batch being processed fails one or more data quality checks the batch status will be DQE complete. To view the data quality errors, click the View Errors button.

The Batch Details page lists the details for the batch including the status for each data quality check. To view the errors for the rule, click the View Errors button.

The Error Page lists the errors for the data quality check.

BATCH MANAGER

Overview

The Batch manager screen is used to track your batches as they are processed through ETL. From this screen, you can see the statuses of your batches as they are processed and review the results of ETL processing, including any data errors or warnings that were encountered during the ETL process.

Default Display

The Batch manager screen appears as follows:

Sort Order

The batches displayed on the Batch Manager screen are always sorted by their associated ID (which is assigned when the batch is created) in reverse order. As a result, the most recently created batches appear at the beginning of the list, while the oldest batches appear at the end. Batches can also be sorted by Batch Type, Comments, Modified Time, Batch Status, and Data Status by clicking on the column headers.

Batch Processing Cycle

From the time that a batch is created until all ETL processing is completed, a batch moves through 4 distinct phases: **Ready To Process**, **In Queue**, **Processing**, and **Complete**.

Ready to Process
Ready To Process is the initial status for newly created batches. Batches in this status are either waiting for ETL resources to become available or waiting for another batch in the current district to complete.

In Queue
Batches are placed into a queue when the system determines that no other batches are running for the current district. This status indicates that your batch is in the system queue, and will be processed as soon as ETL resources are available. You will only see batches in this status when the system is being heavily used or you create multiple batches in a short period of time. Under normal system loads, ETL resources are always available, and batches move into **Processing** status very quickly.

Processing
This status indicates that eDM is currently running the batch through the ETL process. Files in the batch are processed sequentially until the last file completes ETL processing.

Complete
This indicates that all of the files in the batch have completed ETL processing.

Failed
This status indicates that there was a failure in the ETL process while it was attempting to load your data. This is a rare occurrence, and when it happens, you should notify your helpdesk or system administrator.

Data Quality Statuses

Once a batch completes, the system informs you of any data quality issues. A batch can have one of the following statuses:

	Data OK – This indicates that no errors or warnings were found during the ETL process as it loaded the data file(s) in the batch
	Data Warnings – This indicates that some of the data in the batch was questionable, but was loaded into the data warehouse anyway. You should review the results of the ETL process, make any necessary changes to the data, and reprocess the data through eDM.
	Data Errors– This indicates that at least some of the data submitted in the batch contained bad records that were not loaded into the data warehouse. As with warnings, you should review the output from the ETL process, make the necessary changes to the data, and reprocess the data through eDM.
	Data Errors– In the event of a failure during the ETL process, the data status for a batch will be set to Plan Failed.

View Errors

To view ETL Errors in the Batch Manager click the View Errors button.

On The Batch Details page click the View Errors button.

Error files will start with name ERROR_XXX. To View Errors, click the View File button.

FILE	RECORDS	LAST MODIFIED DATE	ACTIONS
ERROR_KL_STUDENT_20171021.TAB	3	2017-10-21 14:37	VIEW FILE
ERROR_SOURCEFILE_MSG_20171021.LOG	0	2017-10-21 14:37	VIEW FILE
Load_SCHOLWHS.PROGRAMS_FACT_LP.log	3	2017-10-21 14:37	VIEW FILE
Load_SCHOLWHS.PROGRAMS_FACT_LP_PARAMS.txt	61	2017-10-21 14:37	VIEW FILE

Error File Display – the page displays the errors in the error file

FILE INFORMATION

FILE TYPE:	ETL Generated File
FILENAME:	ERROR_KL_STUDENT_20171021.TAB

DOWNLOAD FILE (ZIPPED)   DOWNLOAD FILE   DONE

FILE CONTENTS

File Preview

Note: For performance reasons, only the first 100 rows of a file is displayed. To view the entire contents of larger files, you must download the file.

SOURCE RECORD NUMBER	STUDENT ID ERROR COUNT	(K) DISTRICT CODE	(K) SCHOOL YEAR DATE	(K) STUDENT ID	(K) PROGRAMS CODE	(K) BEGINNING DATE	STUDENT ID NOT IN STUDENT TABLE
1	1	2251001	2017-06-30	1123899439	PRG104	2017-06-30	X
2	1	2251001	2017-06-30	1127455269	PRG104	2017-06-30	X
4	1	2251001	2017-06-30	1128833829	PRG104	2017-06-30	X

Filtering Batches

The list of batches displayed on the Batch Manager screen can be filtered by the date they were last updated, the batch status, or the batch type. Simply choose a desired date range/status/type combination from the options at the top of the screen and click on the Filter button.

Searching for Batches

The Batch Manager Search Page provides five ways to search for batches:

- **Search By Batch ID** - To perform a search by Batch ID type in a value for Batch ID and click the Search button. The Batch ID can also be a comma separated list of batch ids.

- **Search By Batch Comments** - To perform a search by Batch Comments type in a value for Batch Comments and click the Search button. The search by Batch Comments is a wild card search. This means that type in any part of the Batch Comments and the system will do a partial search.
- **Search By Batch Status** - To perform a search by Batch Status select a batch status, date range, and click the Search button.
- **Search By Template** - Search by template is designed to search for a batch that contains a specific template. To perform the search first select the data collection, then select one or more templates, and then click the search button.
- **Search Hidden Batches** - The Batch Manager by default does not search hidden batches. To include hidden batches in the search results, click the Search Hidden Batches checkbox.

Hiding Batches

If your system administrator has enabled the ability to hide batches, simply select the checkbox to the left of the batch or batches that you wish to hide from the list of displayed batches(s) and click the Hide from list button at the bottom of the page.

Hiding a batch simply removes it from the list of batches that are displayed in the Batch Manager screen, it does not delete the batch from the system.

Make Batch Visible – to make a hidden batch visible, follow these steps:
1. Go to Batch Manager
2. Go to the Search Tab
3. Click the Hidden Batches radio button
4. Perform the Search
5. View the Batch Details
6. On the Batch Details Page, click the Make Batch Visible Button

Viewing Batch Details

To review the status of any batch within Batch Manager, simply click on the magnifying glass to the right of the batch. This will bring you to the Batch Details page, which contains all of the high-level relevant information for the batch, including the files contained in the batch, the associated comments (if any), the batch ID, and the last time the batch was updated by either a user or the system.

The batch comments can be edited after a batch has been processed. Comments are optional, but can be very useful when trying to locate a batch that has been processed. Comments are limited to 255 characters.

Cloning Files – cloning files copy the files and puts them back in File Manager. This option is useful if the user wants to reload some of the files from the batch and supplement that batch with new or updated files.

Viewing Batch Processing Results

While a batch has an associated data status, the files within a batch also have individual data statuses. To review the results of ETL processing for an individual file, click on the magnifying glass to the right of the file on the Batch Details page. This will take you to the File Details page, but there will be additional information available for the file. This information is generated as part of the ETL process.

The ETL Information page has the following sections:

- **ETL Information** – has when the ETL plan started and ended execution
- **File Statistics** – the information about the records in the file. How many records were inserted, rejected, or no change (meaning the data already existed in the target table)
- **Table Statistics** – for the target table, how many records were inserted or updated

- **ETL Generated Files** – artifacts from the ETL process including:
  - **ERROR Source File Message** – any formatting errors in the file
  - **Load** .log – log file for the ETL process
  - **Params** .txt – parameter file for the ETL process

GENERAL INFORMATION   VALIDATION INFORMATION   ETL INFORMATION

──────────── ETL INFORMATION ────────────

Status:	Plan Complete
ETL Plan Start:	2017-10-20 15:47
ETL Plan End:	2017-10-20 15:47

──────────── FILE STATISTICS ────────────

ROWS SOURCE NO CHANGE :	0
TOTAL RECORDS REJECTED :	0

──────────── TABLE STATISTICS ────────────

RECORDS INSERTED - SCHOOL_ENROLL :	40
ROWS UPDATED - SCHOOL_ENROLL :	0
ROWS DELETED – DELETE UTILITY :	0

──────────── ETL GENERATED FILES ────────────

FILE	RECORDS	LAST MODIFIED DATE	ACTIONS
ERROR_SOURCEFILE_MSG_20171020.LOG	0	2017-10-20 15:47	VIEW FILE
Load_SCHOLWHS.SCHOOL_ENROLL_LP.log	3	2017-10-20 15:47	VIEW FILE
Load_SCHOLWHS.SCHOOL_ENROLL_LP_PARAMS.txt	53	2017-10-20 15:47	VIEW FILE

DOWNLOAD ALL ETL GENERATED FILES   DONE