**IM Data Conversion Validation tool**

**For District/Campus Distributions**

The IM DC Validation Tool for District/Campus Distributions (Package) is an SSIS package under source control.

An extension to this tool for Teacher/Student Distributions is planned for the future.

On your local computer, you will need to have the following folders set up:

E:\DataConversion\IM

A sub-folder for each client should be added (e.g., E:\DataConversion\IM\Hampton\)

This folder will contain the results from the validations. Be aware that the files are date stamped – if the validation is run more than once on a given day, the file will be overwritten.

In VSO/Visual Studio, access this package under HayesIntegration\DEV\DataConversion

Right Click the DataConversion folder and select **Get Latest Version** – this ensures the copy on your local machine is the latest.

Under HayesIntegration\DEV\DataConversion\DataConversion, double click the solution (DataConversion.sln). All SSIS packages for Data Conversion should now show up in your Solution Explorer. Double-Click the ***IMConversion\_DistrictCampusDistributions.dtsx*** to open the package in your work area

The package is now on your LOCAL SERVER. Any changes you make to the package WILL NOT affect the source version control UNLESS they are checked in. HOWEVER, if *Get Latest Version* is clicked AFTER you have made changes to your local package, your changes will be lost. Changes made to run for a specific client do NOT need to be checked in. Changes made to the code or package itself DO need to be checked in. Any changes to the source should be coordinated with your team.

**NOTES on SSIS**

Tasks are grouped into containers

Grayed out containers/tasks are disable – right click on the package or task to enable/disable.

Disabled tasks inside of a container will not process when the container is run.

Containers/Tasks are run in Debug mode

To run a container or task, right click on it and select Execute Container or Execute Task – respectively

Running Containers/Tasks turn yellow then green upon completion

If the Container/Task turns red, there was a problem and it failed – you can review the error(s) in the “Progress” tab – the “Progress” tab is called “Execution Results” when the package in not in Debug.

To end Debug mode, click on the link that says “Package execution completed. Click here to switch to design mode, or select Stop Debugging from the Debug menu.”

**NOTES on IM Data Conversion**

Types of IM Conversion are:

**Biblio** – a conversion from a TipWIN client. Supposed to be super easy cause it is mapping

**Campus Distribution –** sets up books in inventory and district storage and distributes copies to campuses – this package is for this type – also known as “Inventory Only”

**Student/Teacher Distribution –** An extension of Campus Distribution, textbooks are also distribute to teachers and/or students

**Probably more I don’t know about!**

Each TipWeb database will have a list of books (tblMasterBooks) that consists of all available textbooks for that state – when the list posted by the state is available and accurate

A list of Publishers is provided in tblMasterPublishers. Kathie Guthrie is in charge of keeping the data in these tables up to date.

Data Conversion may have two separate spreadsheets for Campus and District (aka Warehouse), or only one spreadsheet for the campus. NOTE: It is also possible that the conversion is only for District books and contains no campus count information)

While the application allows for campus-only books, the rule for conversions is that any books on the campus distribution list must also be district list

Historically, ISBNs were either 10 or 13 characters long. When 13 characters, they would be the 10 character version preceded by 978. These rules are still valid “most of the time” – short ISBNs may require preceding zeros (lost by Excel) in order to match 10 characters. Validation scripts may require closer review than just adding the zeros, however, so you don’t end up with ISBNs such as “0000000N/A”.

ALL alpha characters in the ISBN MUST BE UPPER CASE. If they are not, they will have problems during the audit because the scanner does not equate ‘x’ to ‘X’

“KEEP” field : The validation is designed to be run multiple times – and each time “Import” is reset to yes (1). If manual cleanup has been done, set KEEP to 1 and when the import is reset to 1

**SLC**

The SLC is a four digit code used in Texas that classifies a book. The first digit is the grade (K=0 through 8, and 90, 91, etc. for high school. The second code is the area of study (0=English, 1=Math, etc) and the last two digits are disciplines within that area of study.

If the DC is NOT from Texas, they may or may not be familiar with SLCs. In the application, the SLC field is limited to 4 characters. If the client insists that they need the information in the SLC field, it can be imported into either Notes or Miscellaneous. The data in the UDF field of tblBookInventory shows up in the Miscellaneous field of the application.

**Duplicate ISBNs**

The most common reason for table insert failure is duplicate ISBNs. An ISBN can be added into tblBookInventory only once. Special care should be taken to ensure that no duplicate ISBNs are in the importable data on the District side. Some areas to check for duplication:

**BinNum -** if the client is using bins, it is possible that the books are stored in more than one bin location. Move the bin data from the duplication to the bin2 or bin3 field and delete the duplicate.

**SLC** - If two valid looking SLCs exist in the field, confer with the client on which one is correct.

If all things are equal except the SLC is missing in one or more records, delete the duplicate missing the SLC

If the data is clearly not a code, move the information in the fields to Notes or Miscellaneous and delete the duplicate

**Title/Publisher -** Check for Typos. Look up the ISBN in Master Titles. Google the ISBN to determine the correct Publisher.

**Price** – Use the highest or lowest listed price for all.

**Grade** – Often duplication due to grade is because the ISBN is in there once, for example, for grade 5 and then again for grade 6. Combine the grade to 5-6

**Package Setup**

With the package open in the SSIS work area, review the variables. If you cannot see the variables window, click on the work space, and select “Variables” from the SSIS menu. The following are available for update:

**ConvDB** – This is the DB where the raw data from the client’s SS has been loaded. Standard naming convention is “TipWeb\_IM\_<clientname>”

**StagingDB** – This is the replica of the TIPWeb database. Standard naming convention is “TipWeb\_StagingIM\_<clientname>”

**Server** – This is where the two databases reside. It is currently set to HayesConversion. If working on another server, or your local computer, change the name of the server.

**RejectFilePath** – This is where the reject files will be written to. The standard format is   
“e:\DataConversion\IM\<clientname>” on your local computer

Once the variables are updated, click on the “Show all Variables” icon in the Variables window (second from the right)

**Package Containers and Tasks**

1. **Initialize Tables – sets the data up for processing**
   1. Drop Create Tables – Sets up the current, standard tables (\_DC\_DistrictIM and \_DC\_CampusIM) in the ConvDB database. Don’t run this if table have already been created/populated, unless you want to start over. Disable in the container
   2. Import Campus – This is a data import from the spreadsheet to the standard tables in the ConvDB and is currently disabled. Use SQL import to populate. You may wish to create a copy of the import for do-overs
   3. Pad ISBNs – this is likely going to be used, but as indicated earlier, the ISBNs that do not meet the 10/13 character validation will need to be reviewed
   4. The remainder of the tasks should be run as part of the container (make sure to Disable Drop/Create tables first!)
2. **Pre-Validation** – these are standalone processes that may or may not be needed. Do not run this as a container
   1. Add to Publisher Lookup – Takes each distinct publisher and adds it to the \_PublisherLookup database on HayesConversion. After which the \_PublisherLookup database need to be reviewed to map the LookupPublisher field to the returned Publisher field. This task is a work in progress with the goal of standardizing Publishers names in the database. It should be run ONLY once for each client (unless the raw data changes)
   2. Special Characters – Queries that allow you to research issue that arise when data fields contain invisible, special characters (e.g., CR/LF) See if these queries are helpful if data seems to have problems, but you’ve pulled your hair out for 3 hours and still can’t find them.
3. **Standalone Cleanup Scripts** – >>>>> Each of these tasks is designed to be run individually and outside the context of the package as the results need to be reviewed. The expectation is that these are run ONE TIME, unless the raw data changes. If the script has this line: ***set @StagDB = ?,*** replace the ? with the name of the Staging/TipWeb database (in single quotes) before running in the Query Analyzer.
   1. Cleanup Campus IDs – CampusIDs must match what is in tblCampuses. Leading zeros may have been dropped by Excel, or other formatting issues exist (special characters in the raw data, for example.
   2. Find ISBN w/Multiple Bins – The ISBN can only be inserted into tblBookInventory once, but can be stored in multiple bins. Insert scripts will handle the bins – this script shows you what must be taken apart.
   3. READ ME – Populate Book Bins – instructions inside. Only relevant if district uses bins.
   4. Review Subject Area – Subject areas should be Math, Science etc. Review the distinct subject areas to make sure they make sense. Clean up duplication caused by typos
   5. Invalid Quantity Distributed – Make sure quantities are numeric
   6. Check Distinct Grades – Are Grades Consistent (1 vs 1st)? Clean up – and also a reminder to check for the ISBN/Grade combo that may create distinct records with duplicate grades
   7. Review and Update Material Type – Is the data in this field reasonable for what it should be? Clean up duplicates caused by misspellings.
   8. Review SLC – Is this a code that will fit into the context expected by the application? See “SLC” above
   9. READ ME – Cleanup Duplication – Query looks for any ISBN that can be duplicated based on criteria from above “Duplicate ISBNs”

>>> – Reset Bits before Validating

Resets all “Import” to yes, all Reject Message to null, except when KEEP is turned on

1. District Validation – Run as a Container. This performs all validation and creates a reject file in the reject folder
2. Campus Validation - Run as a Container. This performs all validation and creates a reject file in the reject folder
   1. RUN ME AS SELECT – ISBN not in District Table – a check for ISBNs not contained in the district table. Queries exist to either add (which will require district validation to be rerun) or exclude from import
3. PRE – Table Insert – Run as a container: Each of these tasks should be run ONLY ONCE on the staging/TipWeb database.
   1. PREQUEL to insert tblBookInventory – another check for ISBN duplication in the District data

>>> -- FINAL CHECK – BEFORE Table Inserts

Use these to make sure all required seed data is populated and that there are absolutely, positively no duplicate ISBNs

1. Table Inserts – Run a Container : if all data is clean, these tasks should all run green – and you are done.
   1. Make sure all the tables listed as a notation on the workspace are populated.
   2. Validate that the application works by spot checking
   3. Remind the tester that the reject files will show those files not added
   4. Remind the tester that some data points may have changed formatting. For example, Hampton had Campus ID ‘0020’ and ‘020’ – they were consolidated to both be ‘020’ but in order for counts for this site to match, data for both campus IDs needs to be considered.