ATTAINS Secret Sauce Training – Handout #3: Batch Upload

How is this Useful? Batch upload is one of the more powerful features in ATTAINS. It can also be the most challenging feature to work with. Batch upload allows you to add, modify, or delete just about any piece of information in ATTAINS. Because ATTAINS must ensure that batch upload doesn't fundamentally break your data, it can sometimes be challenging to navigate the errors it generates or even the 'lack of errors' that you sometimes get. The errors that ATTAINS catches are often straight forward to fix. It's the errors that we didn't catch that become problematic. This typically leaves the user having to figure out what's wrong with their files. This can feel like finding a needle in a haystack. These tricks can help you find those errors in a systematic way.

One File at a Time

Occasionally when you are uploading data you may get the 'uncaught error' (as seen below):

> ⚠ Assessment upload could not be completed. No assessments were saved from the files provided.

This error occurs when an error slips past ATTAINS 'front-line' of error checking, but then when ATTAINS goes to save the record, the database data integrity rules stop the record from saving. A big challenge with this error is that as a user, you have no idea what's causing the error, and if you have multiple files that you're trying to upload, you don't even know which file has the problem. It only takes one thing to be wrong to get this kind of error.

An important first step to resolve this is to upload files one at a time. As you move through files one at a time you will quickly discover which file has the error which will allow you to look more specifically at that file. When uploading the files, upload them in this order:

1. Assessments
2. Uses
3. Assessment Types (if you have them)
4. Assessment Methods (if you have them)
5. Parameters
6. Seasons (if you have them)
7. Associated Actions
8. Sources

When doing 'one-at-a-time' uploads, you will likely need to do an 'Update' as opposed to a 'Replace.' The file that you get the 'uncaught error' in will be the one that has the error in it. Use the steps below to try to find some of the common issues with batch upload.

Domains / Case Sensitivity

Ensuring that you're using the correct ATTAINS Domains (or allowed values) is the first thing you should check. The ATTAINS team provides a spreadsheet that contains the current list of Domains for ATTAINS (including any Domains that you are managing as a user). This spreadsheet is available at: https://www.epa.gov/sites/production/files/2020-03/domains_0.xlsx. To keep up to date, you do not need to re-download the file. If you have more recent versions of Excel, you can simply hit the 'Refresh

All' button on the 'Data' Tab within Excel (See Figure 1), and the spreadsheet will update all the domains automatically for you (In order to Refresh, you must first make the document a 'trusted' document).
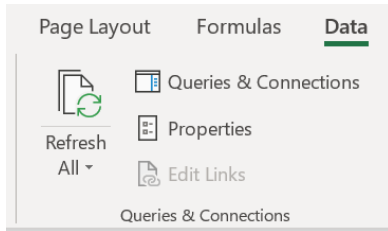

Figure 1. Refreshing Domains in Domains Spreadsheet

A recent update to ATTAINS removed the case sensitivity errors that had been causing issues in the past. However, it is still important to pay attention to Case. To most databases, 'Dissolved Oxygen' is different from 'DISSOLVED OXYGEN' (except Access, it doesn't care, which will sometimes make it hard to find these). ATTAINS does require the Cases for Domains to be consistent *within* your files and *between* your files. For example, you can't have 'DISSOLVED OXYGEN' in your Parameters file, and then 'Dissolved Oxygen' in your Sources file. These are easy errors to make, especially if you're typing information in by hand.

Finding Orphan Records

ATTAINS is expecting comma separated files (.csv). Since you're providing several files as .csv files there is no way for you to ensure that you have all of the records in one file that may be needed in another. CSV files on their own don't have any relational integrity between the files. ATTAINS, on the other hand, requires this relational integrity in order to be able to load data. For example, you can't have a record in your Parameters table that references an Assessment Unit/Use combination that isn't in your Uses file. Since CSVs and Excel don't have the ability to enforce this referential integrity, you will need to use different software. Microsoft Access is a tool that the ATTAINS team often uses to check these relationships. You can quickly link .csv files into Access and then use queries to help you find missing records.

This factsheet is not intended to be a full Access tutorial, and so cannot provide step-by-step instructions for how to do this. However, some basic concepts that will help you include:

- Linking external tables into Access
- Creating Queries
- Joining Tables with Left and Right Joins

When joining the tables, you can follow the similar hierarchy as described above in the 'One File at a Time' section. You'll want to join your files as follows, with the <- indicating the direction of the join, and the data element(s) to join by in '()':

When linking the 'Parameters' table into Access, you will want to change the table name to something like 'Parms'. The word 'Parameter' is a key word in Access and this may cause odd errors when you try to create queries.

- Assessments<-Uses (ASSESMENT_UNIT_ID-ASSESSMENT_UNIT_ID): Looking for any Use records where you don't have an Assessment record

- Uses<-Assessment Methods (ASSESSMENT_UNIT_ID-ASSESSMENT_UNIT_ID and USE_NAME-USE_NAME): Looking for any methods records that don't have corresponding uses
- Uses<-Assessment Types (ASSESSMENT_UNIT_ID-ASSESSMENT_UNIT_ID and USE_NAME-USE_NAME): Looking for any assessment type records that don't have corresponding uses
- Uses<-Parameters (ASSESSMENT_UNIT_ID-ASSESSMENT_UNIT_ID and USE_NAME-PARAM_USE_NAME): Looking for any parameter records that don't have corresponding uses
- Parameters<-Sources (ASSESSMENT_UNIT_ID-ASSESSMENT_UNIT_ID and PARAM_NAME-SOURCE_PARAM_NAME): Looking for any source records that don't have corresponding parameters
- Parameters<-Associated Actions (ASSESSMENT_UNIT_ID-ASSESSMENT_UNIT_ID and PARAM_NAME-ACTION_PARAM_NAME): Looking for any actions records that don't have corresponding parameters

Access is also a great tool for evaluating your domain values, finding parameters that have more than one parameter status (by using Access's Group By and Count functions), and looking for duplicate records.

The Parameters File

We are finding that the parameters file is the one that users are having the most challenges with. A lot of these issues are because this one file has a one-to-many relationship built in. Some rows will duplicate as you have different associated uses associated with the same parameter. This has been a big source of 'uncaught' errors in ATTAINS. You should only have one row for each unique AU/Parameter/Use combination, but, all the columns (with the exception of the use name and parameter attainment code) MUST BE IDENTICAL for each unique AU/Parameter combination. For example, with the two below examples, one is correct and the other is incorrect:

ASSESSMENT_UNIT_ID	PARAM_NAME	PARAM_USE_NAME	PARAM_STATUS_NAME	PARAM_ATTAINMENT_CODE	PARAM_YEAR_LISTED
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Warmwater Permanent Fish Life Propagation Waters	Cause	Not meeting criteria	2014
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Limited Contact Recreation Waters	Cause	Meeting criteria	2014
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Immersion Recreation Waters	Cause	Not meeting criteria	2014

Figure 2. This upload will work. Note that the all the data are identical with the exception of the PARAM_USE_NAME and the PARAM_ATTAINMENT_CODE

ASSESSMENT_UNIT_ID	PARAM_NAME	PARAM_USE_NAME	PARAM_STATUS_NAME	PARAM_ATTAINMENT_CODE	PARAM_YEAR_LISTED
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Warmwater Permanent Fish Life Propagation Waters	Cause	Not meeting criteria	2014
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Limited Contact Recreation Waters	Meeting Criteria	Meeting criteria	
BB_BA-L-FREEMAN_01	CHLOROPHYLL-A	Immersion Recreation Waters	Cause	Not meeting criteria	2016

Figure 3. This upload will NOT work. Note the highlighted values in PARAM_STATUS_NAME and PARAM_YEAR_LISTED. These CONNOT be different.