

Bulk Load Prep Toolbox v1.0

The bulk load prep toolbox contains custom Python tools that are intended to increase the efficiency of data preparation prior to bulk loading into Biotics

Toolbox name: BulkLoadPrepToolbox.pyt

Toolbox access: <https://github.com/PNHP/DataManagement/tree/master/BulkLoading>

Version 1.0 release date: 01/02/2019

Author: Molly Moore, Pennsylvania Natural Heritage Program

1.0 Setting Up the BulkLoadPrepToolbox in ArcMap

1. If you are familiar with Github, clone/download the PNHP/DataManagement repository to your local machine. Use ArcMap Catalog to open the folder where you cloned the repository. The BulkLoadPrepToolbox will be in the folder and ready to use. If you are not familiar with Github, use the following steps to set up the toolbox.
2. Open a web browser and go to:
<https://raw.githubusercontent.com/PNHP/DataManagement/master/BulkLoading/BulkLoadPreBulkLoad.pyt>
3. Right-click on the text in the web browser and click 'Select All'. Right-click again and select 'Copy'.
4. Open ArcMap.
5. In your Catalog, expand the 'Toolboxes' folder and right-click on 'My Toolboxes', select 'New', and click on 'Python Toolbox'.
6. Rename the toolbox, 'BulkLoadPrepToolbox'.
7. Right-click on the newly created toolbox and click 'Edit'.
8. Right-click on the text in the notepad and click 'Select All'. Right-click again and select 'Paste'.
9. Your toolbox is now ready to use!

2.0 Updating Default Parameters (Optional)

Updating default parameters is not necessary, but can save time if you are continually referencing the same dataset or field name in the tool. Use the following instructions to update the default parameters in the Python code.

1. Right-click on the BulkLoadPrepToolbox and click 'Edit'.
2. Press 'ctrl-f' to open the 'Find' dialog box.
3. To update the default value for a particular parameter, find it under the Parameter column in the table below. Type the corresponding Find Search text into the 'Find what:' space in the dialog box. Click 'Find Next' and update the default path or field name in between the quotations. Click 'Find Next' again to ensure there are no other instances of this parameter to update.
4. Once all desired default parameters have been updated click 'File', 'Save', and exit out of Notepad.

Parameter	Find Search
EO Reps Layer	eo_reps.value
Source Feature Point Layer	eo_sourcept.value
Source Feature Line Layer	eo_sourceLn.value
Source Feature Polygon Layer	eo_sourcepy.value
EO ID Field Name	eo_id_field.value
SF ID Field Name	sf_id_field.value
NHD Flowlines Layer	flowlines.value
Network Dataset	network.value
Snap Distance to Flowline	snap_dist.value

3.0 Separation Distance Analysis Tools

The aquatic and terrestrial separation distance analysis tools are used to prepare feature classes and/or shapefiles for bulk load into Biotics by assigning an existing EOID/SFID or new EO/SF grouping string to observations based on separation distance.

3.1. Data Needed to Run the Aquatic and Terrestrial Separation Distance Analysis Tool

- One or more spatial datasets (feature class or shapefile) of observations intended for bulk load (must be projected into a metric coordinate system). This dataset must include:
 - A field that uniquely identifies species using the same values as the Biotics datasets.
 - A field that designates the species separation distance in kilometers.
 - Optional fields for if you are loading observations that have not yet been buffered by locational uncertainty:
 - Locational uncertainty type - a field that designates the type of locational uncertainty. If 'Estimated', observation will be buffered by the locational uncertainty distance for this analysis.
 - Locational uncertainty distance – a field that designates locational uncertainty distance. The observation will be buffered by this distance for the analysis if the locational uncertainty type is 'Estimated'.
- Existing EO Reps spatial dataset (feature class or shapefile). This dataset must include:
 - EOID field.
 - A field that uniquely identifies species using the same values as the input observation datasets.
- Existing source feature point, line, and polygon spatial datasets (feature classes or shapefiles). These datasets must include:
 - SFID field.
 - A field that uniquely identifies species using the same values as the input observation datasets.

3.2. Additional Data Needed to Run the Aquatic Separation Distance Analysis Tool

- NHD flowline layer
- Network dataset built on NHD flowlines (see Appendix 1.0)
- Optional dam/barrier point layer (must be snapped to flowlines – see Appendix 1.0)
- Snap distance in meters which is the distance to the flowline beyond which observations will not be assigned/grouped.

3.3. Post Separation Distance Analysis QC

After running the aquatic or terrestrial separation distance tools, it is good practice to check the results and resolve any issues prior to continuing data preparation steps for Biotics bulk load.

3.3.1. Source Feature Duplicates

If multiple records fall within the minimum mapping distance to one another, they will be given the same SF_NEW value. If bulk loading these records, the user must decide which spatial feature to include as the representative source feature and the other(s) must be removed. Similarly, if a record(s) falls within the minimum mapping distance of an existing source feature, the user must be aware that a new feature should not be created; instead, this would be entered as an update of an existing source feature.

3.3.2. Records with Multiple SF_ID or EO_ID Values

If a record falls within the separation distance of more than one existing SF or EO, multiple values, separated by commas, will be recorded in the SF_ID and/or EO_ID fields. The user must check records with more than one SF_ID or EO_ID values and decide which SF or EO the record should belong to and manually

update the attribute table to include only one SF_ID or EO_ID. In some cases, the user may want to consider combining the new record, along with the multiple SFs and/or EOs, into a single SF and/or EO.

3.3.3. Separation Barriers QC

It is good practice to manually inspect all EO groups against aerial imagery to ensure there are no obvious separation barriers that would require the records to be placed in separate EO groups. To do this, select all records that have the same EO_ID or EO_NEW value and zoom to those records. Make sure they look reasonable and there are no separation barriers. Do this for each group.

3.3.4. Null Values for SF and EO Groups after Running Aquatic Tool

If a record contains null values for all SF and EO grouping values after running the aquatic separation distance analysis tool, it means that the record was further away from a flowline than the snap distance set by the user in the parameters. Zoom to that record and decide whether it should be included in a group and update manually.

Appendix 1.0 – Preparing Aquatic Network Analysis Data

The aquatic network analysis data only needs to be prepared once and then can be used for any future aquatic separation distance analyses. Store these datasets somewhere that can be accessed by those who will be using the separation distance analysis tools.

NHD Flowlines

The NHD flowlines are a part of the NHD dataset and can be downloaded from <https://www.epa.gov/waterdata/nhdplus-national-hydrography-dataset-plus>

NHD Flowline Network Dataset Preparation

The flowlines used in the Aquatic Separation Distance Analysis Tool must be made into a network dataset prior to being used as input. To create a network dataset, follow these guidelines:

- 1.) Create a new feature dataset (use same projected coordinate system as Biotics data).
- 2.) Add the NHD flowlines layer.
- 3.) Right-click on the feature dataset and click New → Network Dataset...
- 4.) Follow the steps in the Network Dataset Wizard. If you are unsure of the options to select, follow these guidelines:
 - a. Select your flowlines layer when asked to select the feature classes that will participate in the network dataset.
 - b. Select No for 'Do you want to model turns in this network?'
 - c. Select None for 'How would you like to model the elevation of your network features?'
 - d. Select No for 'Do you want to establish driving directions for this network dataset?'
 - e. Select the checkbox for 'Build Service Area Index'.
 - f. Use default settings for the rest of the options.

Dam/Barrier Layer Preparation

The Aquatic Separation Distance Analysis Tool has the option of including a barrier layer that will split assignment groups upstream and downstream of a barrier. Dams, areas of point source pollution, or other barriers to upstream and downstream movement can be included in this layer, but all barriers must be represented as points. The point layer used as barriers must be snapped to the network flowlines prior to use in the tool. To snap the points to the flowlines, take the following steps:

- 1.) Open the Snap (editing) tool in ArcMap Desktop.
- 2.) Include the barriers point layer as your input features ****NOTE THAT THIS TOOL WILL PERMANENTLY CHANGE YOUR ORIGINAL DATA. Create a copy of your original data if you wish to have an unchanged backup.**
- 3.) Include the flowlines as the Snap Environment.
- 4.) Choose 'EDGE' for Type and select the distance at which you wish to snap the features (you may have to try a few different distances to get the most accurate snapping effect).