Annotation Load

Author: Lori Corbani

Created: October 16, 2003

Last Modified: October 16, 2003 14:40

1   Purpose of Document

To describe the Annotation Load product.

2   Overview

The purpose of the Annotation Load is to:

1. To load new or delete/reload Annotation records into MGI Annotation tables.

3   Inputs

1. A tab-delimited file in the following format (no column headings):

Table 1: Input File

Field	Name	Example	Type	Max Length
1*	Accession ID of Vocabulary Term	GO:12345	string	30
2*	MGI ID of Object (for example, a Marker or a Genotype)	MGI:12345	string	30
3*	Reference (J:)	12345	integer	
4*	Evidence Code	IDA	string	5
5	Inferred From	SP:12345	string	255
6	NOT	NOT (or blank)	string	3
7*	Editor	dph	string	30
8*	Date	10/16/2003	string	10
9	Notes		string	255

Fields with "*" require a non-null value. Other fields must be present but can be blank.

4   Outputs

1. 2 BCP files (inputfile name is prepended to each bcp file name):
    - VOC_Annot.bcp
    - VOC_Evidence.bcp
2. Diagnostics file (inputfile.MMDDYYYY.diagnostics) that includes all SQL commands issued by the load.
3. Error file (inputfile.MMDDYYYY.error) that includes any data validation errors.

5   Processing

1. Verify Mode.
    - if mode = new and reference is given: delete existing Annotation records for given Annotation Type/Reference; add new annotations
    - if mode = new and reference is not given: delete all Annotations for Annotation Type
    - if mode = append: add new Annotations records
    - if mode = preview: set "DEBUG" to True
2. Verify Annotation Type Name

    If the verification fails, write the error to the error file and terminate.
3. Verify Reference (if mode = new)

    If the verification fails, write the error to the error file and terminate.
4. Load Evidence Codes and Terms into dictionaries for quick lookup. The option -O is used to direct the load to load annotations to obsolete terms.
5. For each line in the input file:
    1. Verify the Term Accession ID is valid for the given Annotation Type.

        If the verification fails, write the error to the error file and skip the record.
    2. Verify the MGI Accession ID is valid for the given Annotation Type.

        If the verification fails, write the error to the error file and skip the record.
    3. Verify the J: is valid.

        If the verification fails, write the error to the error file and skip the record.
    4. Verify the Evidence Code is valid for the given Annotation Type.

        If the verification fails, write the error to the error file and skip the record.
    5. Verify the Editor is provided.

        If the verification fails, write the error to the error file and skip the record.

6. Verify the Date is provided.

    If the verification fails, use the current date.

7. Create appropriate BCP records.

6. Load the bcp files into the database.

6   Usage

To execute:

Python Program: annotload.py

Parameters:

- -S server
- -D mgd database
- -U user
- -P password file
- -M mode (*new, append* or *preview*)
- -I input file
- -A annotation type name ("*GO/Marker*", "*PhenoSlim/Genotype*")
- -R reference (J:####, used when mode = *new*)
- -O load annotations to obsolete terms (default is to NOT load them)