Monthly Data Integrity Checks
Last updated April 3, 2018

NOTES FOR IMPORTING FILES IN NALT/TESA:
1. Spanish taxonomic names should be loaded onto English side
2. Spanish terms should be loaded into English side and only after loading it, Spanish DF/SN should be loaded on Spanish side. See at end of the document for more details.
3. Before loading Ups on Spanish side, make sure there are Spanish DE (English DE should have Spanish DE), otherwise it would not work.

Purpose:
1) To identify orphans on ENG/ES side
2) To identify those terms which need translations
3) To identify those terms which need Definitions translated
4) To identify those terms which need Scope Notes translated
5) To identify those terms which need SC designation
6) To identify those terms which need Definition Source added
7) To identify those terms which are missing BT designation
8) To identify DE without a RK (taxonomic rank)
9) To identify DE with two RK's
10) To add UP's for Taxonomic category (load file on Spanish side)…make sure Pref ES are there.
11) Run report on SKOS file to find oblique quotes/hyphens/apostrophe etc.

Before beginning work in MultiTes, make sure you are using
a) The correct th2 file, e.g. NALT2018.th2
b) That you are on the English side of the database.

It is good practice to note the name of the files you create in this process so you can recall them when reusing the file for further data manipulations.

Procedures:
1) To identify orphans
    a. Report Generator
        i. Select Term Range: Skip terms containing brackets
        ii. Do without the above to remove Spanish orphans
        iii. Term Selection: Orphans only
        iv. Output to: Text File
        v. Run orphan report by clicking on button "Print/Export". Name the file and change to mtm2007\export\(edition year). Click Save. It will reply with "File saved". Close report generator.
        vi. Go to file, search and resolve orphans BEFORE going on to other data integrity checks.

b. RERUN orphan report by clicking on button "Print/Export" (default is "screen" so you should have a "blank" report).

2) To identify those terms, with NO ES (No Spanish translation) which will need translations (and their SN and DF may also need translation):
   a. Report Generator – No ES Report
      i. Select Term Range: Skip terms containing brackets
      ii. Term Selection: Preferred terms only; Approved only
      iii. Display Items: SN, DF
      iv. Select by Relationship: DO NOT have ES
      v. Output to: Text File, clicking on button "Print/Export". Name the file and change to mtm2007\export\(edition year). Click Save. It will reply with "File saved". Close report generator.
   b. Scientific Names in File (no Geo name)
      i. Run report same way but select SC U, save ANSI text file
      ii. Open Excel, paste list in column A and B
      iii. Save file as Text (tab delimited) say save.. yes…OK… .No
      iv. Open tab delimited text or (open as copy) in WordPerfect
      v. Use ALT-F3 to "reveal codes".
      vi. Find [[HRt] and Replace with [HRt] [HRt]
      vii. Click on Replace ALL, count the number of terms
      viii. Under EDIT click on "Find and Replace". Find [Left Tab] and Replace [HRt]ES:<space>. Click on Replace ALL
      ix. Save File as ANSI Windows Text, e.g.
      x. Go to MultiTes, File, Import LoadSciNames.txt on ENGLISH side
         As
         Musca
         ES: Musca

         Feline
         ES: Feline

      xi. If errors from error report, correct
   c. Non-Scientific terms with their SN and DF which need translation
      i. Report Generator – (Remove SC: U)
         1. Select Term Range – Skip terms containing brackets
         2. Term Selection – Preferred Terms only
         3. Display Items – DF, SN
         4. Select by Relationship – DO NOT have ES
         5. Select by Categories-Term should NOT belong to any of the following category---Select SC:U
      ii. Output to: text file. click on button "Print/Export". Name the file and save file to S:\nal\docs\thes\2017. Click Save. It will reply with "File saved
      iii. Email file to Sujata so she can send out for translation.
3) To identify those terms which do not have BT

a. Report Generator
    i. Select Term Range – Skip terms containing brackets
    ii. Term Selection – Preferred Terms only; Approved terms
    iii. Select by Relationship –Does NOT have BT
    iv. Output to: text file. click on button "Print/Export". Name the file and change to mtm2007\export\(edition year). Click Save. It will reply with "File saved". Close report generator.
b. Edit list, removing all the17 top terms that are easy to spot since they are in capitals (e.g., "Biological Sciences"). Send list of terms lacking BT to Thesaurus staff for feedback.
c. Edit terms to add BT in MultiTes.
d. Rerun report to make sure you got them all

4) To identify those terms which need SC designation
    a. Report Generator
        i. Select Term Range – Skip terms containing brackets
        ii. Term Selection – Preferred Terms only; Approved terms
        iii. Select by Relationship – DO NOT have SC
        iv. Output to: text file. click on button "Print/Export". Name the file and change to mtm2007\export\(edition year). Click Save. It will reply with "File saved". Close report generator.
    b. Edit terms to add SC to create ASCII load file. To find SC, search term in MultiTes, open word block, open Multilevel Hierarchy. Note the upper level hierarchy (e.g., Biological Sciences). Click on "Biological Sciences" to reveal its SC, which is "S" in this example. Note that a term may have more than one SC due to polyhierarchy.
    c. Load ASCII file with terms and their SC.
    d. Rerun report to make sure you got them all

5) To identify those terms which need DS designation
    a. Report Generator
        i. Select Term Range – Skip terms containing brackets
        ii. Term Selection – Preferred Terms only; Approved terms
        iii. Select by Relationship – Has DF; DO NOT have DS and vice versa
        iv. Output to: text file. click on button "Print/Export". Name the file and change to mtm2007\export\(edition year). Click Save. It will reply with "File saved". Close report generator.
    b. Edit terms to add DS
    c. Rerun report to make sure you got them all
    d. Also check for Pref terms have USE; MR
    e. Non-pref terms have SC/BT/UF etc.
    f. Pref term has RK, do not belong to U

6) Also check terms that needs RK

7) Spanish terms should be loaded into English side and only after loading it, Spanish DF/SN should be loaded on Spanish side.

Format for loading Spanish descriptors on English side:

Term
ES:
Space

Term
ES:
Space

Format for loading Spanish DF/SN on Spanish side, Spanish terms should be in NALT before loading

Term: Spanish
DF: Spanish
Space

Term: Spanish
DF: Spanish