**GEOREFERENCING 101**

After [setting localities](http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Procedures#Setting_Localities) and [extracting the processed file](http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Procedures#Extracting_.2F_Saving_File), the Excel spreadsheet produced by DEA 2.0 contains two new worksheets: *Main* and *Localities*. The *Localities* worksheet holds all of the specimen records that were previously skipped for not having a matching locality name in DEA 2.0. For these specimen records, new locality names must be created along with their corresponding locality information. This information includes geopolitical units that contain the collecting locality, [properly formed locality name text](http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Procedures#Setting_Localities), and [WGS 84](http://en.wikipedia.org/wiki/World_Geodetic_System) coordinate information.

The overall goal of georeferencing is to locate (as accurately as possible) where a certain specimen was collected.

Helpful Pages:

* DEA 1 <http://hymfiles.biosci.ohio-state.edu/DEA/index.html>
* Google <http://www.google.com/>
* DEA help <http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Help>
* Gazetteers <http://osuc.osu.edu/osucWiki/Geographic_Resources_/_Gazetteers>
* Procedures <http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Procedures>
* Statoids <http://www.statoids.com/>
* State/Province codes <http://osuc.osu.edu/osucWiki/State_/_Province_Codes_for_Countries>
* Google Translate <http://translate.google.com>
* Google Earth
* <http://www.earthpoint.us/TownshipsSearchByDescription.aspx>

A locality name must include any field codes (e.g. T45, CAR01-345, MA-02A-45, etc.), generalized locality terms (e.g. across the road, downstream, well nr. road, etc.), and habitat information (e.g. nothofagus forest, rainforest, sand dunes, etc.). Specific biological associations related to potential host/parasite animals (e.g. feeding on cow, emerged from Nezara sp., etc.) and plant hosts (on flower of lily, from Zea mays, etc.) are omitted from the locality name but included within a separate section within the Main worksheet.

Formatting has been established to:



**[Town/National Park/Reserve], [coordinates on label], [elevation], [field code], [habitat], [random stuff], [political hierarchy]**

**Examples:**

* Label:

S19°09’10,2” W40°01’07,8” BRAZIL: ES,Linhares Res., CVRD, Sta 5, 40km NNE Linhares 16-23.x.1999, semi-decid. for.,”Mata Alta”, MT2, C. Villemant coll.

* Locality:

Linhares Biological Reserve, 19°09'10.2"S 40°01'07.8"W, MT2, semi-deciduous forest / "mata alta", Companhia Vale do Rio Doce (CVRD) station 5, 40km NNE Linhares, ES, Brazil

* Label:

USA: FL. Lee Co., Corkscrew Sec23Twp43sR27E coll. 10.XII emer. 24-25.XII.1987

* Locality:

Corkscrew, T43S R27E S23, Lee Co., FL

**NEED TO KNOWS:**

* When to use / - or commas:

/ used to separate habitats and field codes(if there are more than one)

Ex: ground litter / cloud forest

(-) “or” something

Ex: oak-pine-soybean

, separate different elements **not** habitats

Ex: Pine Creek, Knob River, etc.

* What if a specimen is lab reared?
* Create a locality based on the original collecting information.
* What does stuff like this: Sec23Twp43sR27E mean?
* This refers to the Township information, you can find exact coordinates using this website:

<http://www.earthpoint.us/TownshipsSearchByDescription.aspx>

* Know how to write state highways and US highways
* Google it if you don’t know which it is
* State highway Ex: OH-315
* US highway Ex: US-650
* Common abbreviations:
  + “nr” for near
  + “Twp” for township
  + “km” for kilometers
  + “mi” for miles
  + “m” for meters
  + “ft” for feet
* If the locality is in the USA it ends with the state abbreviation EX: NM, CA, OH
* If the locality is **not**  in the USA, it ends with the full spelling of the country name
* Do not add a space between things like: 8m, 10mi, 20km, etc.
* The “elevation” columns should be in **meters only.**

**Procedure:**

1. Check DEA1 to see what is in the database currently. Since the database is updated regularly it is possible that the locality may have been recently added to the database! Many times this is not the case; however, a similar locality can be used as a good starting point. The database can also be used to reference formatting, but acknowledge that the format of some localities in the database may be incorrect.

* Check for any spelling errors on the label (a quick Google search of the locality usually works). Also, you can and should refer to the specimen if necessary.
* On rare occurrences the correct locality may already be in the database, but needs to be updated (e.g. a locality that uses an alternate name on the label). In this case, use the locality name that’s already in the database. Then, highlight the row (I use a dull red color) in the locality sheet and write “update: [insert what needs changed here]” in the comments column. Then at the bottom of the sheet after the specimen data rows highlight a row using the same color and write something like “update specimen check comments”. This gives your highlights meaning and allows future you and others to determine what the highlights mean.

1. For almost all cases, a gazetteer will be used to accurately determine the coordinates of a geographic feature. Gazetteers can be found on the Wiki page (<http://osuc.osu.edu/osucWiki/Data_Entry_Assistant_%28DEA%29_Help>). Each Gazetteer has a description for which localities to use it for. Most are split by region (e.g. USGS-GNIS for US localities or INEGI for localities in Mexico). Gazetteers are used for verifying the authenticity of the name, identifying the geopolitical hierarchy in which the locality resides, and obtaining the coordinates for the locality.

- Most used Gazetteers

- USGS-GNIS for United State localities

- CGNDB for Canadian localities

- Worldwide localities with no specific gazetteer (check gazetteer page first) use GEOnet

- INEGI for localities in Mexico (in Spanish)

- Lantmäteriet for localities in Sweden (in Swedish)

**Using INEGI for localities in Mexico (in Spanish)**

- Uncheck all “clave” boxes

- Check all “nombre” boxes

- Check “Censo de Población…” box (for population numbers)

- Select estado (state) from first drop down menu

- Click buscar button at bottom

- First drop down box select “nombre de localidad” (name of locality)

- Second drop down box select “contiene” (contains)

- Third type in name of city/what you’re looking for

- Click “agregar” to add (“eliminar” removes)

- Click “cerrar” to close

- Then click “consultar” to generate search

Results:

- The locality will **LIKELY** be urban (ÁMBITO:U) with a decent population > 2000

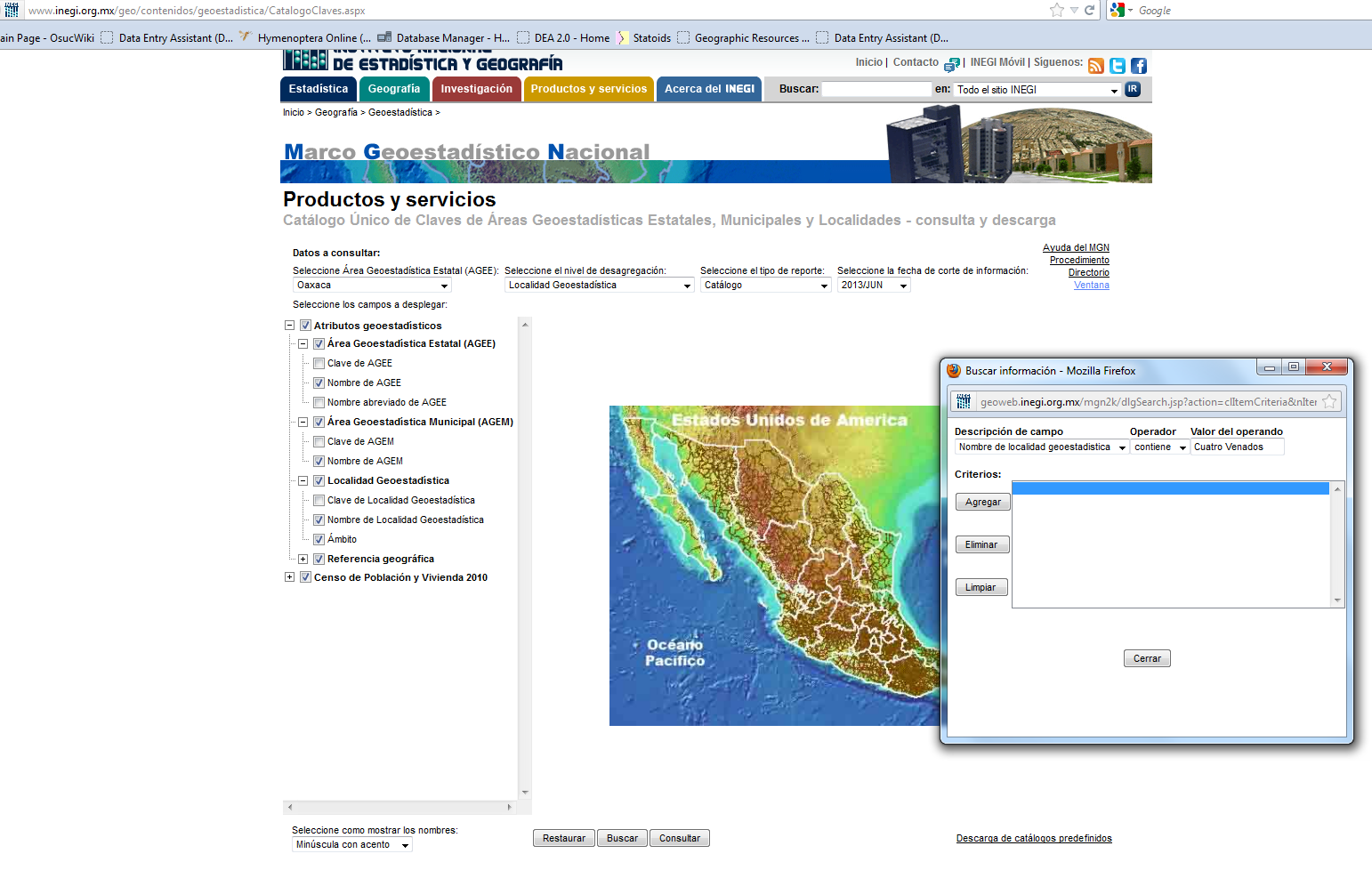
- Use other clues from the label to narrow the search; such as elevation or places the collector was around the time the specimen was collected (check HOL), if available.

- All of the coordinates are given as North and West (you’ll need to change the “O”(Oeste) to a “W” to perform a Google earth search

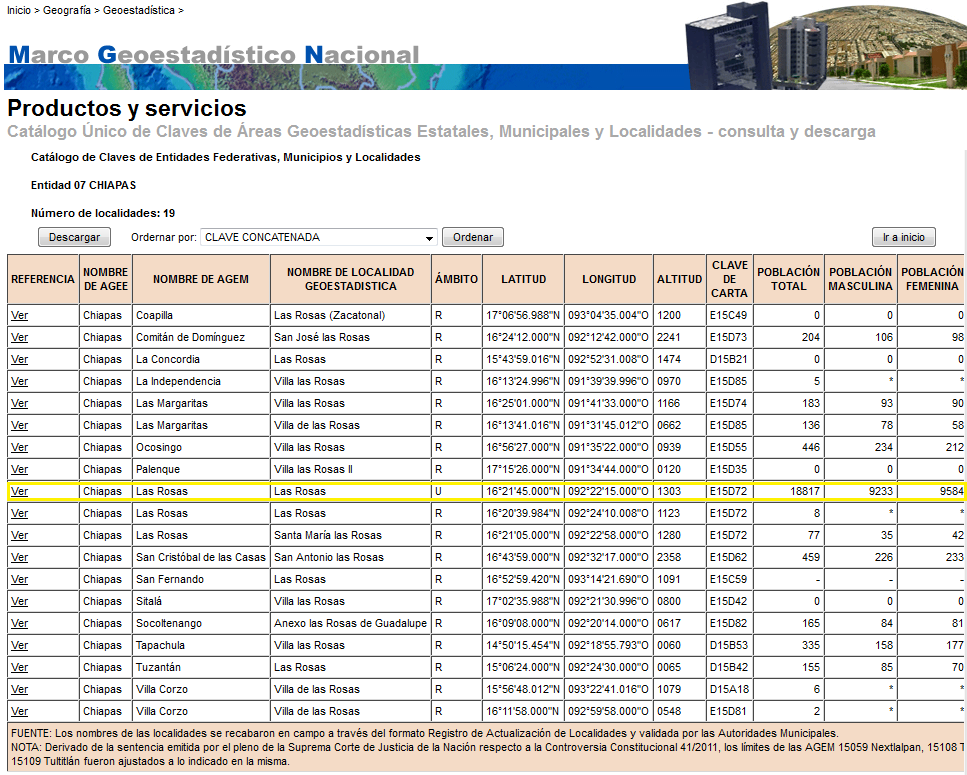
- Don’t use the back button on INEGI

- “Ir a inicio” button will take you back to the search specs

- “restaurar” resets everything



**Figure 1** shows an example of finding a locality using INEGI with the recommended settings applied.



**Figure 2** shows the results for an INEGI search for “Las Rosas” with the appropriate locality highlighted. Notice how the marked locality is urban with a large population compared to the other results.

**Using Lantmäteriet for localities in Sweden (in Swedish)**

- Select län (county) from first drop down box

- Type in name of locality

- Click “sök” to search

- “Bebyggelse” is settlements (town)

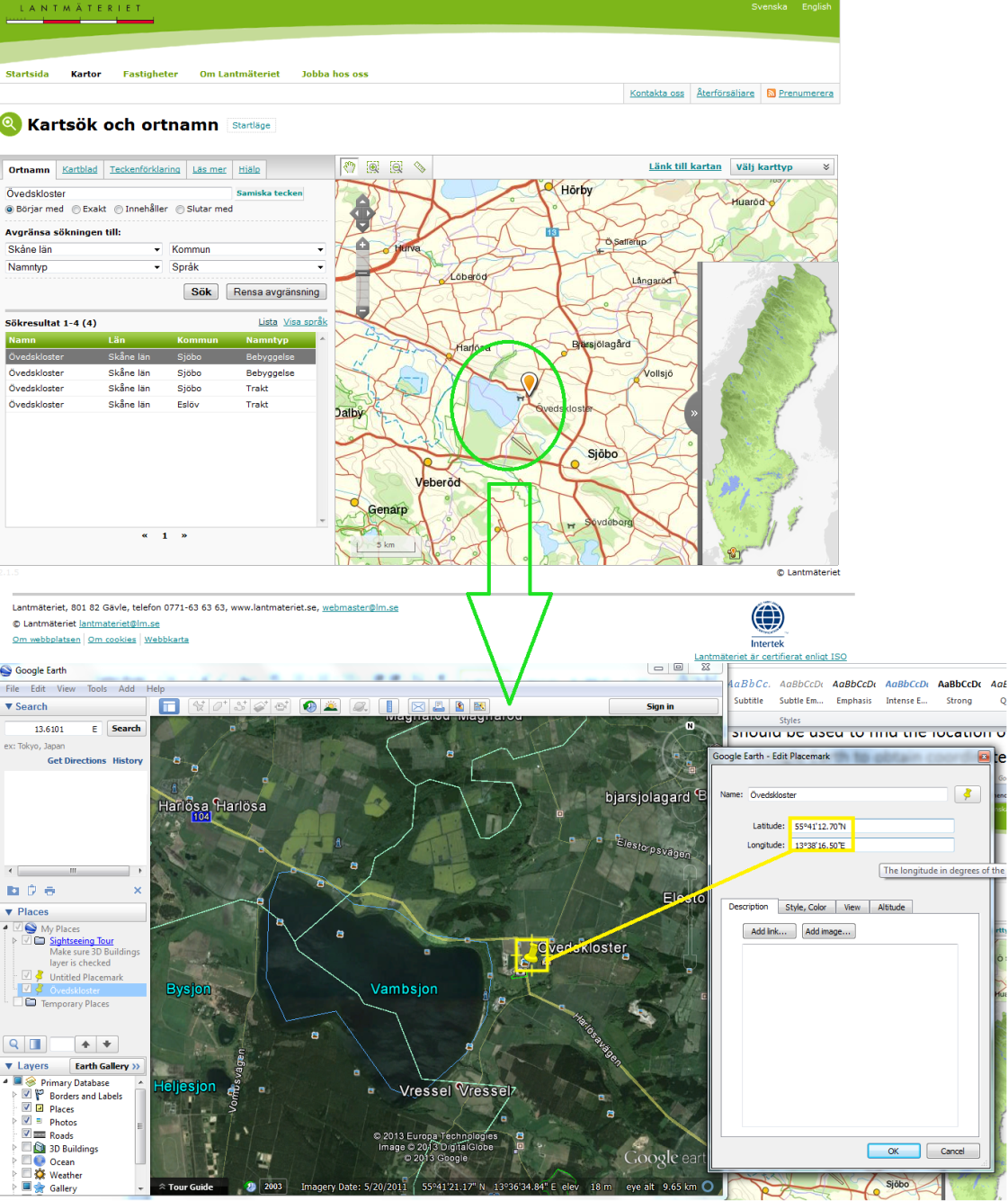
- “Trakt” is a region/area/district

- “Naturskydd” is a nature preserve

- “Kyrka” is a church

- Google translate can be used to decipher the other locality distinctions

- Lantmäteriet does not offer coordinates. It should be used to find the location of the locality. Then, try to find that exact location on Google Earth to obtain coordinates.



**Figure 3** shows an example of finding a locality in Lantmäteriet.

1. After the coordinates for a locality are found and the locality name is formatted (see format above), the rest of the locality information must be filled out.

**- Geopolitical Hierarchy**

When the locality information is gathered for a locality, the geopolitical hierarchy also must be obtained for the locality. The reference used for political divisions is www.Statoids.com or type the parent division (i.e. country) into Hymenoptera Online and browse the subordinate divisions for the accepted spelling. Regardless of the spelling of the geographic division in GEOnet or on the label, use the spelling used in Statoids or Hymenoptera Online. The political division type is the English interpretation of the type unless the division type is in a Romance (French, Spanish, etc.) language (e.g. Indonesian Propinsi -> Province, Ecuadorian Provincia -> Provincia). The Statoids site also provides information on the history of the divisions and alternate divisions which aids in deciphering the current political division for the locality. If a locality is not unequivocally located within a given division, do not use the uncertain division in the geopolitical hierarchy. After obtaining the locality coordinates, the coordinates may be entered into Google Earth to discover the current political division for the point (available for some countries). Only towns or equivalent can be used in the place column, thus townships and similar 3rd level divisions should not be part of the geographic hierarchy.

**-** **The latitude and longitude** for the locality can come from the label, a geographic gazetteer, the internet, literature, personal communication with collector, or *Google Earth*. A coordinate column must be in the format *DD MM SS* where *DD* are the degrees, *MM* are the minutes, and *SS* are the seconds. If there is a decimal, same rules apply. For Example: 43° 17.5’ would be entered as “43 17.5”

**- Latitude direction** (N or S)

**- Longitude direction** (E or W)

**- loc\_prec** is the precision of the coordinates

- POINT (small margin of error) if the known specified area of the locality is less than 325 square miles (18 mi x 18 mi square)

- POLYGON (large margin of error) if larger than 325 square miles (Countries (usually), counties (usually), some forests etc.)

**- Source** is where the coordinates came from (gazetteer, label, Google Earth)

- If location is based on coordinates from a gazetteer, but found on Google Earth. Then Google Earth is the source and in the comments write: “derived from (gazetteer)”. For example: If the label is “1mi N of New Concord” you would use USGS-GNIS to find New Concord. Then, use Google Earths measuring tools to travel ~1 mile north (a little less due to error in exact road tracking) along a major road (if possible). Use the coordinates from Google Earth and in the comments write “derived from USGS-GNIS”.

**- Elevation** is always inputted in meters (only entered if specified on specimen label)

- Use Google to convert feet to meters. Search “xx ft in meters”.

- If a range is given, the first number goes in elevation, the second goes in max\_elevation.

**- loc\_comments** is where you would put any information that helps link the locality to the label. Record any discrepancies, concerns, and/or rationales within locality\_comments column of the localities worksheet as well. For example if something on the label is spelled wrong, the locality has different names, or the coordinates are derived from a different source.

- Examples of comments:

- derived from [gazetteer]/WWW

- xxy is a variant name/alternate spelling/short name/misspelling of xxx

- update: [updates to locality that is already in DB]

- “[gazetteer] coords. adjusted to place over land” (if gazetteer puts coordinates in middle of body of water).

1. **Georeferencing Tips**
   1. Place names with qualifiers that do not occur as the first element of a locality name may have their qualifier name abbreviated (e.g. *Kruger National Park, South Africa*; *Skukuza, Kruger N.P., Mpumalanga Prov., South Africa*)
      1. Recreational Area/Mountains/etc. is abbreviated if not at the beginning of the locality
      2. Spelled out if at beginning
      3. Roads/Avenues/Drives are part of the name of the feature, and therefore should be spelled out.
   2. Populated places in USGS-GNIS & GEOnet are Towns
   3. T.R. Section & RN (Swedish) coordinates do exist and Joe will take care of them
   4. If a feature or place has an English equivalent, use the English equivalent as long as the qualifier is not part of the formal name (e.g. *Parque Nacional Henri Pittier* -> *Henri Pittier National Park*, Rio Pisque -> Pisque River, *Cerro de la Equis* -> *Equis Hill*, *Wadi Saluki* -> *Wadi Saluki* (no English equivalent for *wadi*)).
   5. U.S. localities should end with their appropriate state codes (CA, OH, ID, etc.) (not end with United States)
   6. Political division types should be abbreviated if an appropriate abbreviation is available (e.g. Province -> Prov., State -> St., Município -> Mpio., etc.).
   7. Brazil, Autralia, Canada, and Mexico use division abbreviations found at <http://osuc.osu.edu/osucWiki/State_/_Province_Codes_for_Countries> (from the wiki)
   8. Don’t use statoids for places in Japan – Wikipedia is usually better
   9. Cities in east Asia mean something very specific (we don’t treat them as towns)
   10. Directions from localities should be abbreviated (e.g. N of Papallacta, 4km N Sálakos)
   11. Prefecture is state level in Japan
   12. East Pakistan is Bangladesh
   13. W.T. is Washington Territory and should be entered as WA (Washington)
   14. Sylvania Cal is located in Nevada (already in DB)
   15. Cols. O is an old abbreviation for Columbus, OH
   16. Any uncompleted locality should be highlighted, then add a row at the bottom of the file using the same color highlight to describe the problem. E.g. “can't find anything on Lebang Hara, except that it may be somewhere in West Borneo” and also cite any relevant links that may help in fully determining the locality. Uncertainties can also be pointed out using this system.
   17. Lowland/ mid-elevation/etc. forest is a qualifier making the term one habitat entity
   18. Use the “savannah” spelling of the word (vs. savanna) for localities.
   19. Guano, unless specified (e.g. seagull guano), is a habitat rather than association.
   20. “Mouse nest” is an association (specific enough, most mice are contained within one genus). “Nest” would be a habitat.
   21. Barrenando is formally defined as “drilling”, but can be thought of burrowing. As in “barrenando en caña de azucar” which translates to “burrowing in sugar cane”.