Occurrence Data Entry Protocol

Emma Menchions

2023-11-02

Data Entry

Data Entry

Setting up

1. Navigate to the hackathon google drive link
2. Download the files from the folder "for everyone to download"
    - Place sheets in folder where you can easily access them –> e.g. on desktop "Hackathon_2023-11-04"
3. Open the folder for your assigned group
    - Download and open the empty data entry excel template (Computer screen 1).
        - place in same easily accessible folder
4. Download and open the first assigned field journal pdf file that contains pages in your assigned range (see below) (these contain digitized scans of Harvey's field notes) (computer screen 2 - if available)
    - the assigned pages are as follows:
        - group 4 = pages 18-82
        - group 5 = pages 83 - 145
        - group 6 = pages 146-208
5. Open the file "FloraPNW_Checklist.xlsx" on the same computer as the data entry excel sheet
6. On either computer, open an internet browser window (this will be used to look up the names of locations)

Entering data

Begin with the **first page in a journal.**

1. Enter "**pageNum"** (page number) –> e.g 1 . (circled in bottom corner of pages)
2. **Scan the page for observations (species names)** starting at the top of the page. All species names are acceptable, except for entries which have a collection number beside them (or say "coll" indicating a collection).

- This could be in a **set of species names observed at a particular location** (with no collection numbers)
- **OR the name of a species in the description of a collected specimen** - in this case, just make sure that if there are multiple collected specimens from that location, that one of these associated species was not collected. If it was, do not include it as an observation in a row. The occurrence of this species would be documented in the herbarium and uploaded to GBIF.
- To look for what location and date information is associated with each name, he often put the location and a date and then species associated with it BELOW or BESIDE

3. Enter "**numPage" [REQUIRED]** (number on page) as the number of the observation on the page (e.g. the first species name that matches the criteria in step 3 gets"1", the next one "2", etc.)

   - easiest to enter all of the other data for the observation sequentially on the page and then click and drag a sequence of numbers in excel to number each observation
   - enter observations by columns (e.g. if there are two columns of names on a page, enter the first column and then the second)

4. **"vName" [REQUIRED]** = the verbatim name written in the notes

   - this may be an abbreviation (e.g. Stel med)
   - or full species name (could be misspelled, include these spelling mistakes)
   - or a full species name which is outdated, keep this outdated name, we want to know exactly what he wrote so we can be confident in assigning an updates species name

5. **"vSciName" [REQUIRED]** = your estimation of what the full species name is

   - This is where iNaturalist is helpful. Often the taxonomy might be outdated (indicated by a bracketed name in the search results on iNat).
     - start entering first 3 letters of generic name and species name
     - see if it occurs in the region
     - keep trying with different letters to find best match
     - if there are multiple species possible and/ or you can't decipher the name, place a note in the "dataEntryRemarks" column (e.g. "could not decipher taxon name")
     - Once you are reasonably certain of a full name, start typing it into search bar and finish typing by looking at the result. If the result still pops up then its proper spelling and you can copy it to spreadsheet
     - If it is an outdated name, it will often be bracketed beside another name for the same species in the iNaturalist search, can use this as a spelling guide as well
   - Enter your best guess of what the full verbatim name was meant to be
   - there should be no spelling mistakes, but outdated taxon names should still be used here

6. **"Conf" [REQUIRED]** = enter your confidence level in the **taxon name** estimate (low = l, medium = m, high = h)

   - the names with low confidence will be checked over once all of the occurrences have been entered
   - entries with medium confidence may be returned to if there is time

7. **"sciName" [REQUIRED] = full, properly spelled, updated taxon names.**

   - refer to iNaturalist for updates taxon names
   - If subspecies or variety, just include this epithet, and do not include the "ssp." or "var." abbreviations. (e.g. Instead of Lupinus latifolius ssp. subalpinus, put "Lupinus latifolius subalpinus")

8. **"date" [REQUIRED]** = fill in YYYYMMDD

9. **"locality" [REQUIRED]** = string describing specific location. Does not contain country, province information, but includes place name (island, town, city, area, landmark, trail, road) or position relative to to these features. **Go from GENERAL to SPECIFIC and separate names with ";"**. If there are any words you can't make out or are uncertain of, use [brackets] around the estimated word. I*mportant: Use a web search to make ensure entry of proper spelling of place name!*

   USE CAPITAL LETTERS for each word of city/ town/ municipality/ island names

   SPELL OUT ABBREVIATIONS

   - e.g. Mt –> Mount, Rd –> Rd

   SEPARATE PLACE NAMES WITH SEMI-COLON. Do not tack on terms like "area" .

   e.g. "George Hill area" –> "George Hill; area around George Hill"

   Start with island/ town/ area name...

   - e.g. "Mt. Sutil, Galiano" becomes –> "Galiano Island; Mount Sutil"
   - e.g. "Magic lake horse padoc" becomes –> "North Pender Island; Magic Lake; Magic Lake horse padoc"
   - e.g. "fen below Magic Lake" becomes –> "North Pender Island; Magic Lake; fen below lake"
   - e.g. "small island S of prevost island" –> "Prevost Island; small island to south"
   - e.g. "north side of Cowichan Lake" becomes –> "Cowichan; Cowichan Lake; north side of lake"
   - If there is a string of words like 10 km north of Mount Suitl
     - e.g. Galiano Island; Mount Sutil; 10 kilometers north of Mt. Sutil

10. **"Country" [REQUIRED]** = full name of country of collection (either Canada or United States of America)

11. **"stateProvince" [REQUIRED]** = Full name of province/ state of collection (either British Columbia or Washington)

12. **"island" [REQUIRED]** = full name of island (if collected on island)

    - e.g. Galiano Island or Saltspring Island
    - must contain the word "Island" capitalized

13. **"idQualifier"** = identification Qualifier.

    - only fill out if there was a question mark or other note of uncertainty about the taxon identification from HJ
    - basically replacing question marks and brackets with more formal annotations of uncertainty in ID
      - e.g. Festuca (?) rubra or Festuca (rubra) –> enter here as Festuca cf. rubra
      - e.g. Festuca sp ? –> enter here as Festuca cf. sp
      - e.g Festuca rubra –> enter here as cf. Festuca rubra

14. **"county"** = only fill out if on Vancouver Island or Mainland (otherwise it will be assigned depending on the island that it is on)

15. **"occStatus"** = occurrence status (either present of absent)

    - to save time, only enter if there is a note clearly saying "absent" or a species name is crossed out when there is a list of species with no other species name pointed to (indicating an absence and not just a name update)

  - then enter the word "absent"
- otherwise, assume the species was observed (present) and you do not need to fill this in

16. **"habitat" =** string describing the habitat conditions

- you can use commas when listing things
- to start a new sentence use a semicolon ";" rather than a period
- descriptions about type of species growing in area, but do not list associate species, this will be done in the associated taxa and occurrences columns
- If there are any words you can't make out or are uncertain of, use [brackets]? around the estimated word
  - e.g. "Quercus woodland",
  - e.g. "vernal seepage in limestone rock outcropping"
  - e.g. "edges of lake"
  - e.g. "floating in lake"
  - e.g. "ditch"
  - e.g. "aquatic"
  - e.g. "bog"
  - e.g. "marsh"
  - e.g. "[intertidal]? marsh
- For forest type codes like Cw, Fd, write out the full species names: https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/tree-seed/tree-seed-centre/seed-testing/codes

17. **"assCollTaxa"** = associated collected taxa (names of associated species), growing at same locality at same time

- ONLY denote associated taxa names in this field, if there is a record of a specimen collection (either a collection number or a note about "coll" next to a species name) since these won't be included later but are stored in a separate database (herbarium)
- e.g. Stellaria media 2678 –> "Stellaria media"
- use updated and full names of these taxa
- use iNaturalist to decipher names/ abbreviations of associated taxa and for updated names
- if you can't decipher a name, place the estimated name in [brackets] and make a note in the "dataEntryRemarks" column (e.g. cannot decipher one or more associated species names"
- enter multiple names separated by commas

18. **"assCollOcc"** = associated collected occurrences

- the collection numbers associated with associated collected occurrences
- enter multiple numbers separated by commas

19. **"locationRemarks" =** additional information about location. Separate sentences with semicolons

Anything pertaining to the general location and not what has already been said in the habitat field. If there are any words you can't make out or are uncertain of, use [brackets] around the estimated word

- e.g. "poly 15" (indicating geographic polygon for surveys)
- e.g. "slope 15 percent; aspect 240 degrees" (try to spell out degrees and percents)
- e.g. "mostly open water in lake"

20. **"vTaxonRank" [REQUIRED]** = The most specific level of taxonomy provided verbatim. Only use if the verbatim taxon name is a family name or higher. If genus, species or subspecies, then this information will be filled in later automatically

    - e.g. if "Brassicaceae" –> put "family"

21. **"vElevM" = verbatim elevation in meters.** If there is any information about elevation of the site, input into this column, WITHOUT units. It seems like all of HJ's measurements were in meters. On the off chance that they're not, convert to meters before entering.

22. **"vLat" & "vLon"** = verbatim latitude and longitude.

    For some locations, HJ provided coordinate infromation in lat, long coordinates. If this is the case input this info here either as:

    1. decimal degrees –> e.g vLat = "45.678"
    2. degrees, minuntes, seconds –> e.g. "45 67 89" (do not include " symbols)

    Everything will be converted to decimal degrees at a later stage.

23. **"vUTM"** = verbatim UTM coordinates. If provided, enter in the format:

    - "10U 45000 56890"

24. **"vCoordUncM" = verbatim coordinate uncertainty in meters.** If coordinate uncertainty was clearly stated, input this in meters, without the "m" indicating meters.

25. **numPlantsCode = number of plants code.** HJ often used a number code beside taxon names to quickly indicate rough abundance. The key is as follows:

    - "+" = 1 plants
    - "1" = 1-5 plants
    - "2" = 5-25 plants
    - "3" = 25-50 plants
    - "4" = 50-75 plants
    - "5" = 75+ plants

    When entering these codes, enter "+" as 0 and the rest of the numbers as they are stated

26. **"orgQuantity" & "orgQuantityType" = organism quantity and type.** Only use if HJ indicated abundance in some other way other than the number of plants code.

    - e.g. note about particular species being abundant/ persistant/ dominant/ very few at site –> orgQuantity = "abundant"/ "persistant" / "dominant", "very few" and "orgQuantityType = **"qualitative"**
    - e.g. note about there only being one individual –> orgQuantity = "1", orgQuantityType = "**individuals**"
    - e.g. note about percent cover –> orQuantity = "30 percent", orgQuantityType = "percent cover"

27. **"occRemarks"** = occurrence remarks. Anything pertaining to extra information about the observation that hasn't already been entered OR note about other annotations in book for that occurrence such as an asterisk or if the name has been crossed out and replaced with a new one.

    - e.g. "no usual travel off"
    - e.g. "dominant"

- e.g. “very little, tufts”
- e.g. “asterisk next to taxon name”
- e.g. “genus name originally X but crossed out”
- e.g. “uncollected”
- e.g. “mostly Typha at edge”
- e.g. “two large beds”
- e.g. “varigated”
- e.g. “non-flowering”

28. **“phenology”** = if there were any notes about flowering phenology or life stage (either vegetative, flowering, flowering and fruiting, fruiting and flowering, budding, budding and flowering)

    - e.g. “non-flowering” enter as –> “vegetative”
    - e.g. “fl” enter as -> “flowering”
    - e.g. “has fruits” or “has seeds” –> “fruiting”

29. “recordedBy” = people there at time of collection (full names or initials) only enter if there are additional names that Harvey denoted of people with him collecting.

    - e.g. “PJ” = enter as”Pam Janszen”

30. **“idBy” = identification by**. Only fill out with names if there was a note about someone other than Harvey identifying it.

31. **“dataEntryRemaks” =** any observation with remarks here will be checked over. Write whatever notes are important to completing data entry for that row/ what needs to be helped with.

When finished entering:

- Save file as .csv
- Upload to your group’s google drive folder with the following name: “HJ-8_occ-entry_G-[group numer]_v[version number].csv”

Important Notes:

1. **SAVE FREQUENTLY** (command S)

2. If certain names can’t be deciphered or a location/ time can’t be attributed to it, **still include as a row in the template,** with a valid **pageNum** an **numPage** entry. It can be removed later (this will help for re-finding things on the page). To help remember what you had a difficult time deciphering, make a note of it in “dataEntryRemarks” and you can additionally attempt to fill out the information and place square brackets around the uncertain parts –> e.g. [add tophar]?

3. **DON’T enter any taxon observations with either a collection number or the note “coll no #” or “coll”** (this indicates a specimen was collected and therefore, we don’t want to duplicate data already in herbarium databases)

4. **AVOID entering repeated observations of same taxa** at same locality, habitat/ sub location and same date

    - there will be checks to make sure this doesn’t happen during data processing, but best to avoid to save time now, if the repetition is really obvious

Tips for efficiency

1. When you first start a new page or new event (location and date), **count the** number of observations you think you will have. Enter the date and locality info for the first observation and then **copy for the amount of rows you think you will need. The same applies for associated collected occurrences**
2. When it comes to big chunks where most observations are associated with collections - only write down ones that aren't a collection at that locality - be careful it can sometimes be tricky, but you can write down coordinates for the collected specimen as for the occurrence.
    - Additionally, you can copy and paste large lists of associated occurrence numbers and taxa names to multiple rows to save time
3. Using iNaturalist for efficient species type and spelling check
    - Starting to write a name in the search bar on the explore page and then filling the rest of the name out by looking at the result that pops up
    - copy and paste this to sheet

Metadata

- Abundance codes: from page 1 of the HJ-7 field journal
    - "+" = 1
    - "1" = 1-5
    - "2" = 5-25
    - "3" = 25-50"
    - "4" = 50-75
    - "5" = 75+
- Codes like Cw, Fd, Md all refer to forest type codes which can be found here: https://www2.gov.bc.ca/gov/content/industry/forestry/managing-our-forest-resources/tree-seed/tree-seed-centre/seed-testing/codes
- List of common abbreviations:
    - Aira car = Aira caryophyllea
    - Alli acu = Allium acuminatum
    - Alli amp = Allium amplectens
    - Atri pat = Atriplex patula
    - Anth odo = Anthoxanthum odoratum
    - Apha arv = Aphanes arvensis
    - Arct col = Arctostaphylos columbiana
    - Bra arv = Sinapis arvensis (formerly Brassica)
    - Brom ste = Bromus sterilis
    - Brom hor = Bromus hordeaceus
    - Brom sit = Bromus sitchensis
    - Card oli = Cardamine oligosperma

- Care ino = Carex inops
- Cerast arv = Cerastium arvense
- Clar amo = Clarkia amoena
- Clay par = Claytonia perfoliata
- Cyno cri = Cynosurus cristatus
- Cyno ech = Cynosurus echinatus
- Dact glo = Dactylis glomerata
- Dauc pus = Daucus pusillis
- Delp men = Delphinium menziesii
- Distich spi = Distichlis spicata
- Dant cal = Danthonia californica
- Eleo pal = Eleocharis palustris
- Elym gla = Elymus glaucus
- Epil ade = Epilobium adenocaulon?
- Epil pau = Epilobium parviflorum?
- Erod cic = Erodium cicutarium
- Erio lam = Eriophyllum lanatum
- Eryt ore = Erythronium oregonum
- Gali apa = Galium aparine
- Gara can = Rosa canina
- Gera mol = Geranium molle
- holo disc = Holodiscus discolor
- Holc lan = Holcus lanatus
- Hypo rad = Hypochaeris radicata
- Koel cri = Koelaria macrantha
- Lact mur = Lactuca muralis
- Lepi vir = Lepidium virginicum
- Loni his = Lonicera hispidula
- Loma utr = Lomatium utriculatum
- Litho par = Lithophragma parviflorum
- Lych cor = Silene coronaria (formerly Lychnis)
- Meli sub = Melica subulata
- Mimu gut = Erythranthe gutatta (formerly Mimulus)
- Madi gra = Madia gracilis
- Mont fon = Montia fontana
- Mont par = Montia parvifolia
- Nemo par = Nemophila parviflora
- Osmo chi = Osmorhiza berteroi (formerly chilensis)
- Plan mar = Plantago maritima
- Plect con = Plectritis congesta
- Pter aqu = Pteridium aquilinum

- Poa pra = Poa pratensis
- Raco can = Racomitrium canescens
- Rosa gym = Rosa gymnocarpa
- Sani cra = Sanicula crassicaulis
- Sedu lan = Sedum lanceolatum
- Sedu spa = Sedum spathulifolium
- Sela wal = Selaginella wallacei
- Stel med = Stellaria media
- Symp alb = Symphoricarpos albus
- Tori jap = Torilis japonica
- Trif dub = Trifolium dubium
- Trif micro = Trifolium microdon
- Trif miceph = Trifolium microcephalum
- Trif tri = Trifolium willdenovii
- Trit hya = Triteleia hyacinthina
- Vero arv = Veronica arvensis
- Vici hir = Vicia hirsuta
- Vici sat = Vicia sativa
- Vulp bro = Festuca bromoides