**The species inventory metadata capture form**

The goal of this form is to capture key information that is reported by surveyors when theyperform a survey for species. Such reports provide critical information on not only a list of species found, but the taxonomic, geospatial, habitat and temporal scope of the survey, the methodologies employed, and ultimately an assessment of completeness of the survey. The form has been tuned to capture these essential details and is not meant to be an exhaustive capture device. It also attempts to standardize something that is quite heterogeneous in terms of reporting quality, style and content. We expect that many surveys generate **a lot** of blank fields. The best surveys will have many fields filled in.

What role should guessing or assumptions play in the data entry process? In developing this form, we decided to not complicate data entry tasks by asking those entering data to also analyze and infer too much from the reports. This would require an even further level of training and would conflate data capture and downstream analysis in bad ways. We therefore came to agreement that the right approach here is to is to make sure that we treat this form as a verbatim capture device, and make little to no inferences except in some key places (e.g. total area covered, measures of effort, completeness).

**General Information**

**Dataset name**

* Use the title from the study as it was published

**Dataset Identifier**

* For this one we just need to put the name of the place where the study was done, and the taxon. For ex: “South Jordan Herps”
* You should get this one from the Priority Reserve Checklist (that way we can keep track of it)

**Citation Reference**

* This is the citation for the published paper.
* We need to be consistent when entering. So APA style has been chosen to use. 1) the first author has first initials, and then last name, and the coauthors are all formatted "last name, initials" 2) Note the formats on date (e.g. no parentheticals) and the Volume(Issue):Page numbers formatting.
  + Ex: L. Belbin, Daly J., Hirsch T., Hobern D., LaSalle J. 2013. A specialist’s audit of aggregated occurrence records: An ‘aggregator’s’ perspective. ZooKeys 305(1): 67–76.

**Citation identifier (DOI etc.)**

* This is where we provide the DOI, ISSN, ISBN, or URI numbers that reference the survey. Usually one is provided in the study. DOI is the prefered one to use.
* If there is no DOI, etc., provided in the study, then find the number through a google search with the study’s title. Crossref.org or doi.org can also be used to look up the DOI number.
* Sometimes the publication is considered a book or a serial and will have an ISSN or ISBN instead.
* In some rare cases, no identifier can be found and just put “none found”
* We need to use the right syntax for this section, i.e. ISBN:xxxx or DOI:xxxx
  + DOI:10.3897/zookeys.305.5438
  + ISBN: 0553804677
  + ISSN:0036-8075

**Citation URL**

* This is the web address for the study.
  + Ex: http://faculty.washington.edu/leache/wordpress/wp-content/uploads/2011/02/2 005HerpReview.pdf
* This is not expected to necessarily be permanent but provides a mechanism to allow another (quick) link between the metadata, the data and the original source.
* Not always present so use “none found” when it isn’t.

**Metadata recorder first name/Metadata recorder last name**

* You put your first and last name in the appropriate boxes.
* Option of “+” in case more than one person was involved.

**Role**

* **Initially recorded by**
  + Name of person who initially filled out the form.
* **Vetted/edited by** 
  + Name of person who reviews/edit the form.

**Taxonomy**

**Citation reference of taxonomic authority**

* The document that specifies the published taxonomic authority used for species names described in the checklist. Provide full citation in APA format.
* This needs to be cited consistently by using APA style.
* There can be more than one reference and the “+” symbol allows us to add more.

**Habitat and Environment**

**Was information about the habitat(s) surveyed are provided?**

* **Habitats were reported by site**
  + Select this one when the author(s) describes the habitat details for each site surveyed. When you select this option, the following questions show up under the “Site details” section: “Habitat inclusion” and “Habitat exclusion” (allowing you to fill out information for each specific site)
* **Habitats were reported but not by site**
  + This one should be selected when the author(s) provides a description of the habitat surveyed, but doesn’t specify it by site name/number. In this case, “Habitat inclusion” and “Habitat exclusion” will pop up next, and you can provide the general information about the habitat.
* **Habitats were not reported**
  + If this is the case, no information about the habitat surveyed was provided on the paper, and you move on to the next question. (“Habitat inclusion” and “Habitat exclusion” questions stay hidden).

**Habitat inclusion**

* The idea behind the "habitat inclusion" question is to capture what the authors report about the habitats that they sampled
* We'll add a "Habitat inclusion" question under "Site details", and build a controlled vocabulary section later on (so this is something we can start looking out for as we fill out the checklist).

**Habitat exclusion**

* This question is to capture the habitats that the author(s) purposefully excluded during the survey for any reason.
* It should only be filled out if the paper explicitly provides information about the habitats that were not surveyed. For example, the surveyors could have purposely avoided agriculture habitats in their survey.

**Geospatial information**

**Geospatial scope**

* This is the location that the checklist represents (you can usually pull it directly from the title).
* For example, if the checklist is for the “Herps of Acadia National Park”, then the geospatial scope is “Acadia National Park”.

**Total reported area covered by geospatial scope (km2)**

* In the case of the example above, the total area covered by geospatial scope would be the area of Acadia National Park.
* If the author reports the size of their geographic scope, we’ll capture it. If they don’t, we’ll just leave it blank.

**Total area covered by species inventory (km2)**

* The question is meant to capture what is actually inventoried, and is smaller than the area represented in the checklist. For ex: if a study claims to cover “Acadia National Park”, and the authors set up 10 one-ha plots as their inventory, the answer to this question would be 10ha.
* Since our approach here is to avoid making inferences as much as possible, if the author(s) happens to have written the answer to this question, then we can capture it. If this is easy to infer from what is provided by the authors, such as the example above, to then capture it (**note**: this is one of two places we make an exception to our rule about verbatim capture). If this is not easy to infer a simple answer, better to leave it blank.

**Number of separate sites**

* How many separate sites were inventoried? If the author(s) isn’t clear about whether they surveyed one or more different sites, just leave it blank.

**Site details:** If author(s) provide information about each individual site, then we report them in this section.

* **Site name**
  + This is the name provided by the author to define the site and separate it from other sites.
* **Longitude**
  + This needs to be in decimal form. Be careful when entering because the author might state the latitude before the longitude.
  + Use the same converter when converting to decimals: <http://transition.fcc.gov/mb/audio/bickel/DDDMMSS-decimal.html>
* **Latitude**
  + This needs to be in decimal form. Be careful when entering because the author might state the longitude before the latitude.
  + Use the same converter when converting to decimals: <http://transition.fcc.gov/mb/audio/bickel/DDDMMSS-decimal.html>
* **Habitat inclusion and Habitat exclusion questions here when you select “Habitats were reported by site”, as previously mentioned.**
* **“+”** allows you to provide all of the information under “Site details” to as many sites as needed.

**Local Conditions**

**Reported on the ground climatic or weather conditions**

* This would be any reported conditions by the author(s) in their survey.
  + Ex: Flooding, Migration, etc.
* These are conditions that are reported on the ground, i.e. not e.g. climatic means for an area.

**Reported conditional impacts**

* This field is to report any events that may have impacted the survey. They can be positive and negative.
  + Ex: Flooding may have caused less data to be collected.

“+” allows you to make more than one of these sections if you need

**Temporal scope**

**Survey time block(s)**

* Since the author(s) don’t always provide specific dates when the surveys were taken, we decided to keep the year, month and day boxes all separate, to allow for more flexibility.
  + Start year:
  + Start month:
  + Start date:
  + End year:
  + End month:
  + End date:
* Just fill out as much information about the dates as possible.
* “+” allows you to make more than one of these time blocks if you need

**Number of (months, days, hours) spent surveying in each time block**

* This section is to capture the studies that provide a time block and then break it down into less specific time frames.
* For example, if the study was conducted for a month but the author(s) say they only surveyed for 3 days out of the month, the correct information to enter is “3”.
* If a study has more than one survey period, just take the average of the different number of days per time block.

**Temporal units (months, days, hours)**

* This is the unit used by the author(s).

**Start time/End time**

* Here we can fill out the times when the survey started and ended each day (if provided).
* For example: if the researchers surveyed the area from 9AM until 4PM each day.

**Was the study conducted during the day?** (Y/N)

* This should be answered yes only if it is stated by the authors in the study.
  + **Provide the hours the study was conducted during the day as**
    - This pops up only if yes is clicked. Three choices are provided if the authors provide additional information. Each choice brings up a pop up to capture the details that were provided by the author
      * An exact number of hours
      * A range of hours
      * A start and end time

**Was the study conducted during the night?** (Y/N)

* This should be answered yes only if it is stated by the authors in the study.
  + **Provide the hours the study was conducted during the night as**
    - This pops up only if yes is clicked. Three choices are provided if the authors provide additional information. Each choice brings up a pop up to capture the details that were provided by the author
      * An exact number of hours
      * A range of hours
      * A start and end time

**Season(s)**

* No need to try to figure out the season(s) when the study took place from the information given.
* Only fill this one out if the season(s) was specifically mentioned in the paper.
  + Spring
  + Summer
  + Fall
  + Winter
  + Dry
  + Wet

**Organismic information**

**Prospective taxonomic scope**

* The set of species which were searched for.
* The idea here is to check all information that an author give us, and not assume any hierarchy. So if an author reports that they searched for "small mammals", we just check "small mammals", with the knowledge that we don't know exactly what the author considers the taxonomic scope for small mammals.
* If the author reports they search for small mammals and then later mentions they searched for rodents and shrews, we check "small mammals", "rodents" and "shrews". Again basically capturing verbatim what the author reported. If we want to make assumptions later as part of our scripting we can do that, but we don't make assumptions during the entry of the form.
* Here is the list:
  + Mammals
  + Large mammals
  + Medium mammals
  + Small mammals
  + Rodents
  + Bats
  + Primates
  + Birds
  + Herps
  + Reptiles
  + Amphibians
  + Plants
  + Trees
  + Shrubs
  + Herbs
  + Insects
  + Molluscs
  + Other

If you click on “other”, a question should pop up

**Taxonomic scope excluded from the study**

* If the author(s) reports any taxa that were intentionally not searched for.
* Excluded taxon scope should be a subset of prospective taxonomic scope. i.e if the prospective taxonomic scope is “Reptiles” but the authors for some reason says “We did not identify plants” you don’t need to enter that into the excluded taxonomic scope section.

**Was taxonomic completeness reported?**

* No, taxonomic completeness was not reported
  + This is checked if the study did not report the taxonomic completeness
* Yes, reported as incomplete
  + This is checked if they found all but a few species that are known to live in that area.
  + If this is checked then a further question pops up ask “Did reporter provide assessment of how incomplete?”. This is a yes or no question and should be answered yes only if it is stated by the author(s).
* Yes, reported as complete
  + This is check if they found all the species that were suppose to be in that area.
* If the author reported some measure of completeness, use that preferentially as much as you can, but do also consider the below.
* If the author did not report some measure of completeness, use the following rubric If the inventory included survey work and is not a compilation. The measure of completeness we want to capture is an upper and lower bound, with 25 percent increments. We also need to make sure to have a notes field for capturing some content about how data entry folks made their assessment. This should be "light" -- not an extensive description but enough so we can capture the core thinking.
  + How many sites were surveyed across the full geospatial scope? Was coverage across space adequate? Did it cover different habitats?
  + How well did sampling cover time periods? Did it cover different time periods in the season, and during the course of a single day (e.g. day and night)? The more coverage is adequate, the more complete the survey was likely to be.
  + Was effort adequate? Effort adequacy really depends on a lot of factors. 100 trap hours might be great for common small mammals, but not so good for capturing insect diversity.
  + Were abundances or common-rare species noted? If so, you can use information on rare and common species along with effort to make a better estimate of completeness. Lots of reports of single individuals or just two individuals per species is often an indication that more sampling will yield more taxa
  + If the inventory also included compilations/expert knowledge, can you assess effort associated of the compilation efforts. Give more weight to multiple sources of compilation and adjust completeness estimate upward as a result.
  + Were there other factors that contributed to sampling deficiency or completeness? Give a quick explanation in the "completeness assessment notes field".
* If the inventory was ONLY compilation:
  + If the authors report completeness, use that information first.
  + If there is no reported completeness, then... How many different type of resources were consulted? Multiple sources (e.g. Museums AND expert opinion) should yield higher levels of completeness.
  + Within types of sources, how many different items were consulted. For example, how many different museum collections were examined, or how many different local experts. The more items, the higher completeness.
  + Can you assess the effort the authors have put into compilation beyond the type or resources and number of items used to create the compilation? This could be data quality effort, measures of time spent per interview as a finer grain view of effort, etc.

**Developmental stage**

* This section is designed to capture if the survey was conducted by only looking at certain developmental stages.
* This is filled out when the author specifically report development stages.
* This list allows for multiple answers:
  + Adult
  + Juvenile
  + Larval
  + Seedling
  + Vegetative

**Developmental stage excluded**

* This section is designed to capture if the survey was conducted by excluding certain developmental stages.
* This is filled out when the author(s) specifically report excluded developmental stages.
* This list allows for multiple answers:
  + Adult
  + Juvenile
  + Larval
  + Seedling
  + Vegetative

**Distributional status**

* If the author(s) state whether they looked for specific distributions of the species in their study.
* This list allows for multiple answers:
  + Native
  + Non-native
  + Invasive

**Distributional status excluded**

* If the author(s) state whether they looked for specific distributions of the species were excluded from their study.
* This list allows for multiple answers:
  + Native
  + Non-native
  + Invasive

**Growth form**

* This section is for mainly plant inventories.
* If the author(s) report on growth form of the plants that they surveyed in their study.
* This list allows for multiple answers:
  + Tree
  + Shrub

**Growth form excluded**

* This section is for mainly plant inventories.
* If the author(s) report on growth form of the plants that they excluded from their study.
* This list allows for multiple answers:
  + Tree
  + Shrub

**Size class**

* This section is an open field/mutli-lined to capture the size data provided by the author(s) in their survey.
* An example is if the author(s) only surveyed trees that were greater than 10m in height, or fish that were greater than 3cm in length, etc.
* We haven’t been able to define it yet, so we are keeping it the way it is for now until we do more studies.

**Size class excluded**

* This section is also open field/mutli-lined that captures the exclusions of size data by the author(s) in their survey.
* An example is if the authors excluded data from trees that were less than 10m in height, or fish that were less than 3cm in length, etc.
* This section is also not defined yet, so we will keep it the way it is for now until we do more studies.

**Methodology**

**Was compiled data also collected?** (Y/N)

* Choose yes when authors report using a source of information that has been assessed o compiled from elsewhere. Such sources might include museum data, expert knowledge, etc.
  + **Type(s) of compilation**
    - You will see this question on the form if you select “Yes” to the previous question.
    - Expert knowledge
    - Primary literature
    - Museum specimens
    - Other
* Choose no this if the authors did not consult existing compiled or assessed data sources and instead only performed the inventory in the field.
  + Examples:
    - “*A generalized search for amphibians and reptiles was undertaken at this site by a pair of researchers…*”
    - “*Data were collected from random field observations during a number of visits*

**Inventory type**

* The definitions for the choices are here: <https://docs.google.com/spreadsheet/ccc?key=0AhrY0qRdO4budC1mSUNWTDlkOXBVMjcza2Y3aV84SkE&usp=drive_web#gid=2>
* Some loose working definitions of the choices:
  + Restricted Search - Use if the author mentions transects or plots.
  + Open Search - Less restrictive than restricted search - a particular search area is defined, but effort is presumed to be standardized across those areas. “*We searched sites for a length of time proportional to their area*”.
  + Opportunistic Search - This is a search such as often performed in efforts such as eBird where a surveyor goes to an area and does a search that doesn’t have a pre-defined shape and size of the survey effort, nor necessarily a clear effort measurement associated with it. It may still be relatively complete.
  + Trap - Select this one if the authors used stationary traps or nets. Also use if the authors used secondary evidence such as scat.

**Shape of search area**

* This question should only show up if you click on “restricted search” or “open search” for “Inventory type”, because only these two types of searches have a defined shape. The options for shape of search area are: “rectangle”, “circle”, “line” and “polygon”.

**Protocol name**

* In theory we think this is the name of an established protocol, as defined in the protocol reference. In practice we have never run across something that we think would be called a “protocol name”; however we are keeping this field. We believe that this field is useful in the future when we do more studies because some surveys might provide this detail. An example protocol name is “Pollard walk” used in butterfly surveying.

**Protocol detail**

* This field is almost never filled out in practice and if it is the data is very unstructured. We are going to leave it how it is for now until we work on more surveys. This is so we aren’t over specifying up front.
* Don’t have to fill it out unless it is stated or details are provided.
* Usually seen in restricted searches. These searches provide more details about things, like transects, area, plot.
* This section is used if the surveyors change the protocol to fit their needs or there are any variations to the published protocol. For compilations that don't have a standard protocol, or have an unpublished protocol, the description of how goes here.
* No protocol details are need for opportunistic searches.

**Protocol citation**

* If the study cited the methods of another study, this goes here. Note this is not where you would include the taxonomy authority used. This field is for citing methods only.
* Cite in APA format for consistency.

**Protocol reference**

* This is where we provide the DOI, ISSN, ISBN, or URL to link to a referenced protocol, if it is reported. For example: If the surveyor says: “We used the method of Jetz et al. to survey birds”, you’d place the DOI or URL to point to that paper.

**Were abundances reported for species inventoried?** (Y/N)

* Often, in addition to a list of species, author(s) will included the number of individuals of each species. This is abundance information.
* If the author(s) only provides a checklist of species with +/- and not numbers then the answer for this question is no.

**Was there a maximum number counted for abundances?** (Y/N)

* Displayed only if the above question answer is Yes.
* Some methods involve counting individuals, but only up to a certain number. For example, the methods may sate “we counted individuals up to a maximum of 10”. In this case if abundances were reported and a species had 10 individuals recorded, we know they author(s) saw a minimum of 10 individuals but may have seen more.

**What maximum was specified?**

* Use in the case that the answer to the above question is Yes.
* In the above example, the authors counted to a maximum of 10 individuals, so the answer to this question would be “10”.

**Were absences reported?** (Y/N)

* Oftentimes, the researcher(s) may provide information about species that were expected to be found in the surveyed area, but were not observed during the survey.

**Methodology: Effort**

Capturing effort is a central goal of the form. Effort can be reported in many forms, such as Trap-nights, Person-hours, or Net-nights. If multiple effort types are reported we want to capture each of these individually. In this case, use the repeater control to add multiple effort types and enter the information.

**Effort amount:**

* Represents the total effort amount for this particular effort unit. For example, if the authors only used traps as a sampling methodology, and reported 1125 trap-nights, the answer to this question is 1125.

**Effort unit**

* Record the unit that the effort was reported in. In the example above, where the author only used traps and reported 1125 trap-nights of effort, the answer to this question is “Trap-nights”
* At times an author may report a unit that is not in the controlled vocabulary for this question. In this case, pick “Other…”. We should use this field as little as possible, so if you find that you need to pick “Other…” contact the form administrator to see if your effort unit can be added to the controlled vocab for this field.

**Other effort unit**

* Use this to enter the other effort unit if it is not in the controlled vocab. We should use this field as little as possible, so if you find that you need to pick “Other…” contact the form administrator to see if your effort unit can be added to the controlled vocab for this field.

**Was there a more granular breakdown of effort reported?** (Y/N)

* Sometimes a more granular breakdown of effort will be reported. For example, a study may report 20 total person-hours, but further report that 15 person hours were conducted during the day, and 5 at night. Or a study may report 25 sampling days, 10 of which were at site #1, 5 at site #2, and 10 at site #3. If this is the case, answer Yes.

**Granular effort breakdown**

* In above scenario, the answer to this question might be “Day: 15, Night: 5” or “Site #1: 10, Site #2:5, Site #3:10”

**Was effort reported by the author or calculated by the data entry person (Reported/Calculated)**

* Since effort is so important, we may at time break our rule of strict verbatim data entry and calculate effort. An example might be a study in which the author states they surveyed “more than 100 transects”, each transect was “25 minutes”, and they surveyed each transect “up to six times”. In this case, the effort can be reasonably calculated as 100\*25\*6/60 = 250 Person-hours (assuming transects were performed by only one individual).
* If such a calculation was performed by the data entry person, choose “Calculated”

**Description of effort calculation**

* If you chose “Calculated” above, describe how you arrived at your number. Just add something like the description above.

\*You can also add an extra section of effort if there were multiple types of effort

**Methodology: Sampling Strategy**

**Were specimen vouchers taken during the survey?** (Y/N)

* Use this if the whole animal was preserved, or an entire bird skin, etc. This is in contrast to a sample where the only material taken might be a piece of hair, or an ear clipping.

**Were samples from the organisms taken during the survey?** (Y/N)

* Use this if only a small part of the animal was taken, such as hair or an ear clipping. This is in contrast to a voucher where the entire animal is preserved.

**What kind of samples?**

* This one only shows up if you answer “Yes” to the previous question. The answer options are: blood, tissue, fecal, and/or other. If you select “other”, then a box will show up allowing you to specify the collected sample.

**Were measurements made from the organisms taken during the survey?** (Y/N)

* Use this if measurements such as snout length, weight, etc. were taken.

**Recording agent type**: (Human/Machine)

* This is to differentiate the agent that makes a decision to record information. If a camera trap that has an automatic sensor or a proximity sensor is used, it is a machine that makes the “decision” to record presence. If a human writes down the presence of a bird on a sheet of paper, then it is the human that decides. If a human walks around with a recorder and presses “record” when they hear a bird song, the agent would still be be “Human”.

|  |
| --- |
|  |
| **Data quality**  **Quality assurance procedure:**   * Refers to the procedures that were used for quality control on the part of the surveyors either during the survey or after the surveying was completed. Post-survey QA/QC could be checking taxonomic identifications, refining measures of effort, etc. * Since there are no specific instructions for how to fill out this field at this point, we've decided to put it on hold, until we enter data into the checklist. Later, as we start having a better sense of what people report in these studies, we can return to this issue and try to get a better vocabulary for this field. * Ex: inventory could have been reviewed by an expert, * Is this generally a post field process? It could be during or after the survey.   . |
|  |