# 

# BIOSCAN for Flying Insects - UK Sample Metadata SOP V2.0

| ***Version*** | ***Date*** | ***Details*** | ***Contributors*** |
| --- | --- | --- | --- |
| **2.0** | JUNE 2022 | based on Darwin Tree of Life manifest but extensively revised to support BIOSCAN collections. | Lyndall Pereira, Mara Lawniczak, Petra Korlevic, Alex Makunin, Katie Woodcock, Vickie Brookes, and members of the Darwin Tree of Life Sample Working Group |

Correct and comprehensive recording of sample metadata is critical to the long-term utility of the work we do. Metadata will link the sequence data we generate to their origins.

**Please read this Standard Operating Procedure (SOP) in full before completing the** [**Sample Manifest**](https://docs.google.com/spreadsheets/d/1jWl2o6RD4JQiItprXBj0jcWDkQJyUvHxiAd7VWpwV_k/edit?usp=sharing) as it contains detailed guidance on how to record metadata. Please also watch this [**training video**](https://youtu.be/yrZAEg6FgWA) to assist you in filling the manifest in rapidly and correctly.

## 

## OVERVIEW

This document contains column by column instructions for capturing metadata at a single sample level (a sample can be a whole insect in a microtiter plate well, or a part of an insect in a well). To accomplish this, the manifest contains two data entry tabs:

**TAB 1**: “Contributors” records the information for all people who contributed to the batch of samples recorded in the manifest, for example, the collection, identification, and/or preservation of samples.

**TAB 2**: “Metadata Entry” records information on each individual sample/specimen.

There are also two further tabs:

**TAB 3**: “Well IDs (for copy/paste)” contains the well numbers on a plate so this does not need to be manually entered but can be copy/pasted into TAB 2 in the “*TUBE\_OR\_WELL\_ID*” column for each submitted plate. If entering data at plate level only, the correct term can be copy-pasted from column B in this tab (i.e. “PLATE\_ONLY”).

**TAB 4**: “Data Validation” is solely for populating drop down menus and cannot be modified by partners without seeking confirmation from [bioscan.info@sanger.ac.uk](mailto:bioscan.info@sanger.ac.uk). It is possible to see all fixed term options in advance of adding metadata by viewing this tab. This tab has been locked for editing on Google Sheets, however when this is downloaded as an excel file, it is no longer locked. If using Excel, please be careful not to change any terms in this tab otherwise it will affect the drop down menus in TAB 2.

**TAB 5**: “Example Metadata Entry” is for illustrative purposes only and shows an example dataset for 3 plates that represent 3 trap collection events made over 3 months that are ready for shipping to Sanger. To help illustrate potential uses of the manifest, each 96-well plate is highlighted in a different colour and each trap collection event is separated by a black line to demonstrate that a plate may have specimens from more than one collection event. The first plate has plate-level only data and only requires one row in the manifest as all metadata is identical (row highlighted in blue). The second plate (rows highlighted in yellow) and third plate (rows highlighted in green) need to be filled in at well level because metadata are different for some of the specimens. Note that we allow one partially filled plate per shipment batch so that all specimens in the latest collection event can be processed and analysed together.

**Orange** metadata fields (i.e. columns) are mandatory – we must have meaningful information to proceed with sample processing. **Purple** columnsare strongly recommended if the data have been collected, if the information was not collected the entry can be left empty. **White**  columns are not validated and mostly allow descriptive text relating to the specimen that you feel is important to record, we encourage data entry if the data have been collected.

After sequencing, all raw sequence data will be submitted to an open access repository linked to a specimen identifier (SPECIMEN\_ID) which will be assigned to each specimen after manifest validation. You will receive the SPECIMEN\_ID for each specimen in the post-validation manifest as an additional column (see *Section 7* of this document for details on the post-validation manifest).

## 1. Contributions

All people who hold primary responsibility for your samples and anyone who contributes to individual collections at a level warranting authorship to resulting publications should have their names added to “TAB 1 Contributors” in the manifest that accompanies each submission of specimens in plates. These fields are also highlighted in **orange** indicating that they are mandatory.

## 2. Manifest submission guidelines

To fill in the manifest, you must always download or make a copy of the [Google Sheet linked here](https://docs.google.com/spreadsheets/d/1jWl2o6RD4JQiItprXBj0jcWDkQJyUvHxiAd7VWpwV_k/edit?usp=sharing) each time you wish to start a new manifest, this will ensure you are working with the latest version. You can make a copy of the Google Sheet (File > Make a copy) and work there online or offline (File > Make available offline), or download it as an Excel file (File > Download > Microsoft Excel (.xlsx)). It is important to save the manifest with its original formatting (i.e., not as a .txt or .csv file) otherwise you will lose the drop down menu terms. You may have some issues if you are working on the Google Sheet directly in a browser other than Chrome.

Note that if you copy/paste, you will overwrite any data validation and the error flags that pop up to let you know you have entered data incorrectly will no longer occur, so please be careful when you use copy/paste to enter metadata to ensure that your entries meet the requirements (e.g. “*FEMALE”*, not “*F”* or “*fem”*). We recommend first doing a manual entry of several specimens so that you can take advantage of the data validation and identify where there are likely to be issues when using copy/paste.

Before we are able to accept any samples, your completed manifest must go through a “validation” process. When you have completed filling in your manifest, but **prior to shipping** your samples, please send the manifest (as a link to a Google Sheet or as an Excel file) to [bioscan.info@sanger.ac.uk](mailto:bioscan.info@sanger.ac.uk) to go through the validation process. This ensures that all fields are complete and interpretable for every submitted sample. If any issues with the information provided are identified (e.g., missing mandatory entries, duplicate rows, incorrect date formats) the sample manifest will be returned to you along with a list of detected issues and you can resubmit once you have resolved these issues.

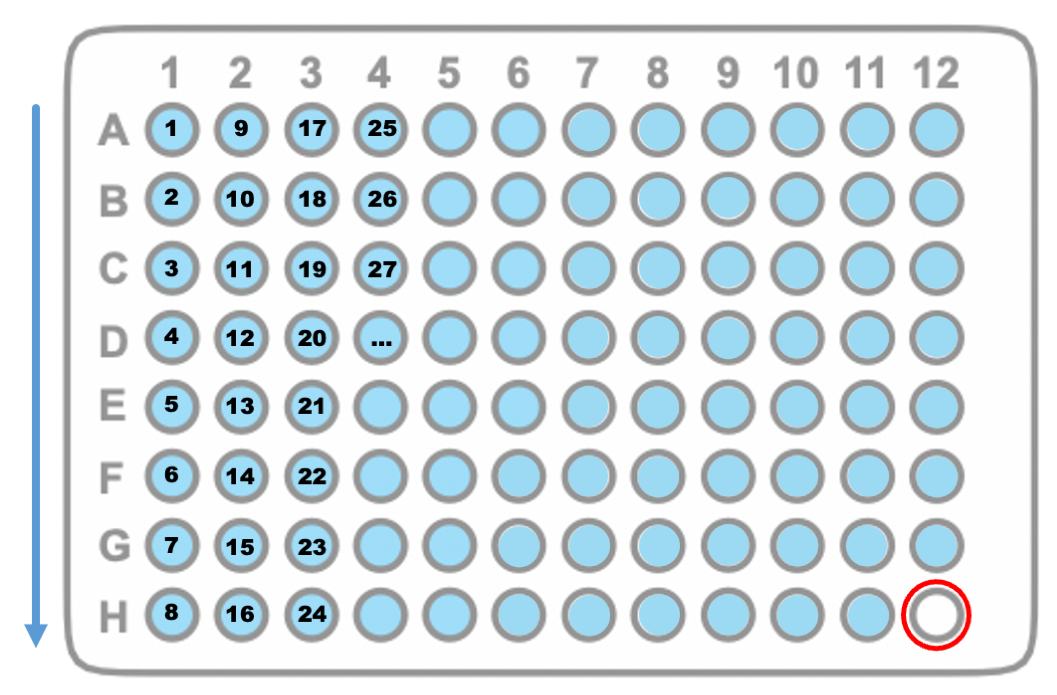
Please carefully read the guidance in this SOP for each field, and attempt to get your submitted manifests as close to the guidance as possible to avoid many iterations of validation.

## 3. Sample submission

We can only accept samples in the 96 well microtiter plates that we provide to you as part of the project. Specimens (in ethanol or lysis buffer) should be shipped in skirted plates (ThermoFisher Scientific P/N AB2800) with strip caps (Fisherbrand Domed 8-Cap Strips P/N 14-230-231) properly sealed with a sealing tool to prevent evaporation.

### 3.1. Plating Specimens/samples

1. Before beginning, ensure your plate is in the correct orientation (e.g. first well is labelled A1). Plate insects in a vertical direction in 96 well plates working down columns (e.g A1-H1 followed by A2-H2 and so on, see figure below). Please ensure that plates are filled entirely with 95 insect specimens and only position H12 (circled in red below) is left empty. In some cases H11 will also be used as a control, these wells will be indicated on specimen plates with a red marking and by using separate caps not connected to the cap-strip so they remain capped during specimen plating. Plates should be sealed with tightly fitting 8-strip caps.



1. Plates can comprise specimens from more than one trap collection event – this is much preferred to submitting partial plates. Please endeavour to fill plates completely. Partial plates are acceptable for the final collection event before a shipment is to be sent to Sanger.
2. Any insects left over after plating can be placed into the corresponding 50 mL Falcon tube with the residual catch ethanol, labelled as described in the CATCH\_LOT field below.
3. Some insects will be too large to be plated in a microtiter plate well and in these cases, leg(s) should be put into the well and the remainder of the specimen should be preserved as deemed appropriate by the partner depending on longer term plans. The body part that has been plated should be recorded in the ORGANISM\_PART column (see below). For large insects (e.g. a bumblebee), place a piece of a single leg in the well. For smaller insects (e.g. a Marmalade hoverfly), place 1-3 legs in the well, preferably all from one side of the organism so that the voucher specimen keeps a representative of each leg type. The remaining tissue that has not been plated can either:
4. remain with the Partner and it will be up to them as to how and where to store them, including linking the specimen (use the VOUCHER\_ID field for this, see *Section 6*, column AG) with the SPECIMEN\_ID (provided after validation of the manifest, see *Section 7*) of the ORGANISM\_PART. This can also be pre-arranged with the sample repository of choice (e.g. a museum).
5. be placed into the corresponding 50 mL Catch Lot tube. These tubes and their contents are returned to Sanger and may be used for R&D studies and not necessarily returned to partners.
6. Every submitted plate should contain exactly 95 samples. **Please always leave the final well (H12) empty of specimens –** this well must contain the PRESERVATIVE\_SOLUTION but no biological material. The only fields that must have entries for the empty well H12 are RACK\_OR\_PLATE\_ID, TUBE\_OR\_WELL\_ID, PRESERVATIVE\_SOLUTION (which should be the same as the rest of plate), and ORGANISM\_PART (select “*NOT\_APPLICABLE*”).

### 3.2. Metadata entry

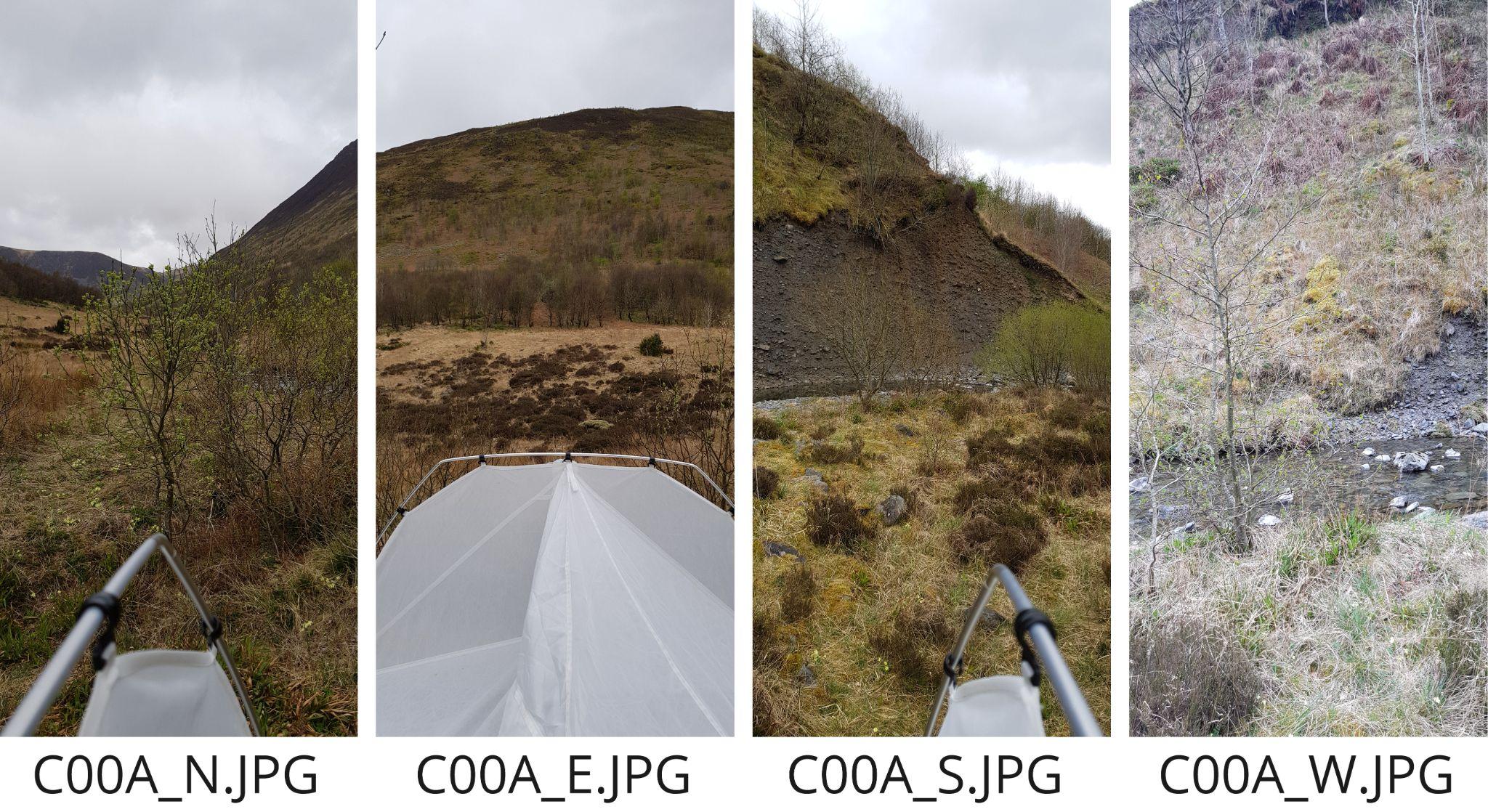
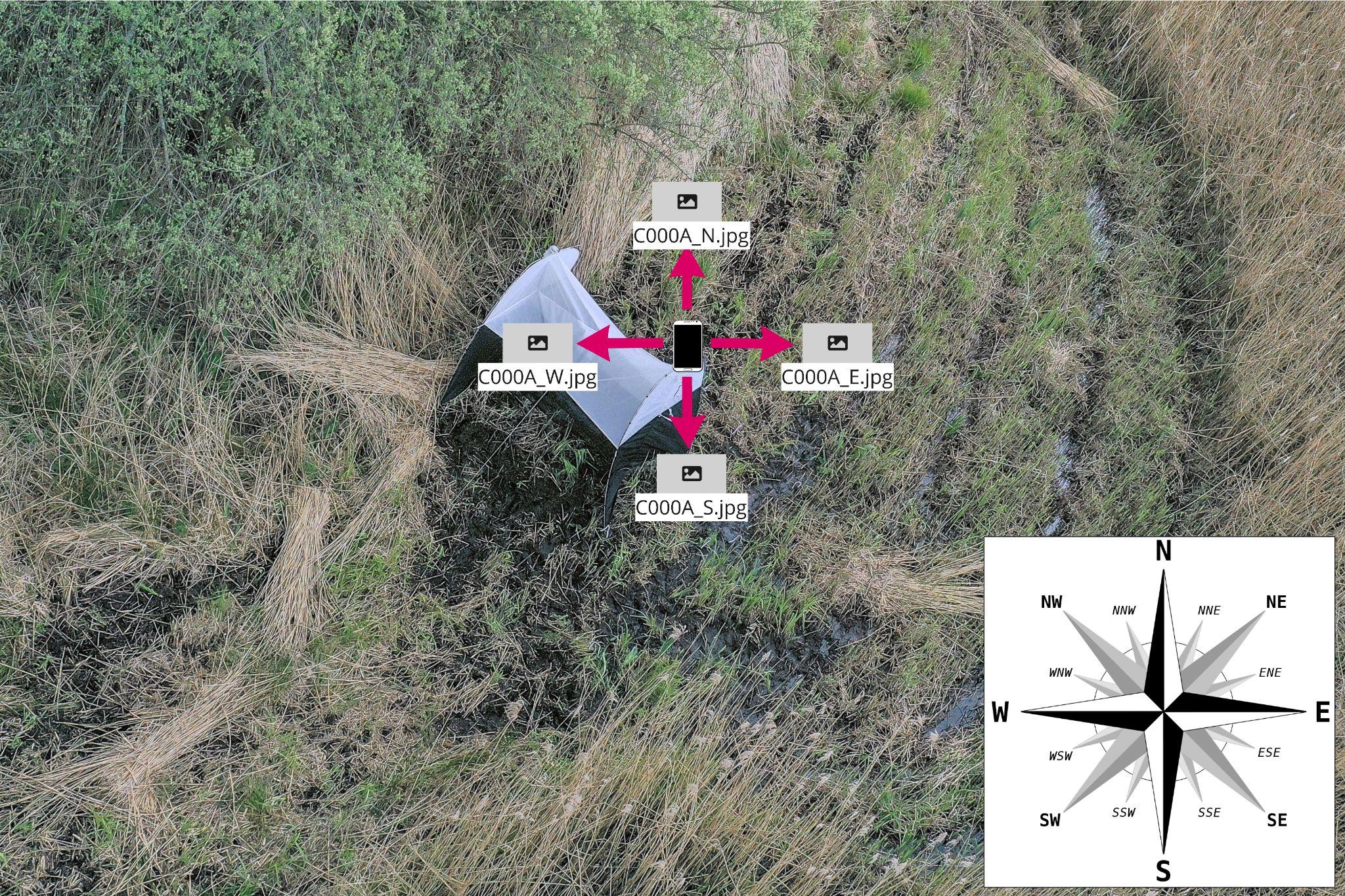
1. If a plate represents one catch lot and all well-level data is identical (i.e. all specimens are WHOLE\_ORGANISM, no specimens have taxonomic identifications, no wells are empty apart from H12, and the entire plate contains samples from a single catch), then this can be submitted in one row of the manifest avoiding the need to copy the same information across 95 rows. If a single entry for an entire plate is possible, then the TUBE\_OR\_WELL\_ID column must be filled with the term “PLATE\_ONLY” to make this clear If any of these conditions are not met, please follow the guidance and submit 96 rows of metadata per plate.
2. The manifest has drop-down menus in some fields. If you discover a missing term in the drop-down menus, new terms can be suggested by contacting us. Please only request new terms if the absence of the term is likely to affect many samples.
3. **Information must be entered for all fields below with orange highlighted names** [in the Google Sheet version of the manifest, these fields are represented by cells with an orange fill. The fill will go white when an entry has been made to help you identify which mandatory fields still require data. All fields with orange highlighted namesare mandatory and must be filled with the appropriate information. For all other fields, if you have collected the information, please add it. If not, you may leave these fields empty. It is permitted to add information for a subset of specimens on a plate.
4. **All dates in the manifest must be formatted consistently as YYYY-MM-DD** (ISO 8601). If only year or year/month are known, use **YYYY** or **YYYY-MM**, respectively.
5. In fields that are “free text” we ask that you use only the core alphanumeric characters, plus full stop “.”, hyphen “-”, underscore “\_” and spaces (summarised in coding parlance as ” **-\_.a-zA-Z0-9**”). Please avoid “|” (the vertical pipe symbol) except where we indicate it should be used to separate elements in a list. Please **do not** use “special characters” (such as other punctuation and “logical” marks: “#”‘;:?!@\*()[]{}/\,=+”, etc.).

**When you have completed the metadata entry and before you send any samples, please email your manifest or send a link to the BIOSCAN team at** [bioscan.info@sanger.ac.uk](mailto:bioscan.info@sanger.ac.uk). Please name your manifest PARTNERCODE-YYMM-VERSION.xls, e.g. PART-2205-V2.xls; PARTNERCODE is a four letter site code assigned during partner onboarding, and YYMM is the year and month of your initial manifest submission to the email address above rather than the timing of the collection or shipment.

**4. A note regarding photographs**

There are three categories of photographs recommended or required for each collection event, 1) Site, 2) Catch and, 3) Specimen Plate with barcode label. Please upload photos after each collection event to the folder containing your partner code at [this link](https://drive.google.com/drive/folders/1HJLUo7_V2lyh31xJA-l0-2qqCK0AoN4V?usp=sharing). Further guidance on the photographs can be found below:

**Site photographs (recommended)**: To capture the habitat and conditions at each Malaise trap per monthly catch period, take four photographs from the top of either trap bottle in the directions of the four cardinal points, starting with North, then rotating the camera 90° in a clockwise direction until you have taken four photographs in the order: North, East, South, West. Take a photo of each cardinal point direction from the top of the malaise trap catch bottle (choose one bottle to do this from for migratory, 2-bottle Malaise traps). Label each photo with the CATCH\_LOT code (either CATCH\_LOT code is fine for migratory, 2-bottle Malaise traps) followed by an underscore (\_) and the cardinal point direction of the photo (e.g. C000A\_N; C000A\_E; C000A\_S; C000A\_W). Upload all site photos in this section.



**Catch photographs (recommended)**: After the 24 hour trapping period, when specimens have been tipped out of the trap bottle and into the sorting tray, please take one photo of the catch in the tray from directly above, please include the 50 mL Catch Lot tube in the photograph so that the label is visible in the photo. Label the photo with CATCH\_LOT code, followed by an underscore and “tray”, e.g. C000A\_tray.



**C000A\_tray.JPG**

**Specimen Plate photograph (required)**: Once the specimens have been plated, please take one photo of each plate from the underside at an angle so that specimens and barcode are both visible. These are useful to us for quality control checks on receipt of the specimen plates. Label the photos with the human readable text on the plate, if you have taken more than one photo then add the suffix “\_ventral” for the underneath photo and “\_lateral” for the side photo of the plate with the barcode label visible.



**BSCN\_000.JPG**

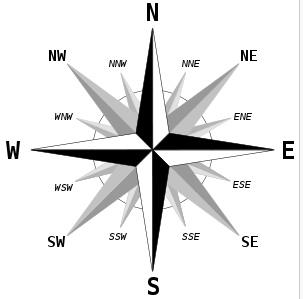
# 5. Tab1: Contributors

All contributors to the sample collection and the overall study should have their names and affiliations listed in this tab. Please list all people who contributed at a level warranting authorship to samples contained within TAB 2. This may include for example people who are responsible for the collection, identification, preservation of samples or have overall responsibility in the project.

1. **SURNAME**: First letter capitalised, rest lower case (e.g., Darwin)
2. **FIRST\_NAME**: First letter capitalised, rest lower case (e.g., Charles). If middle initials should be included, please add these here (e.g. Charles R. ).
3. **PRIMARY\_AFFILIATION**: Add the primary affiliation for each individual. For those without an official affiliation (e.g. volunteers), partners should discuss the use of their affiliation as the work was done on their behalf. Additional affiliations and complete addresses will be requested at the time of publication by emailing everyone on the Contributors Tab.
4. **EMAIL\_ADDRESS**: Add a reliable email address as this is the primary route by which we will contact you for authorship queries.
5. **CONTRIBUTION:** List the contribution(s) made by the individual (e.g., collected specimens). Please list “Primary Contact” as an entry here for the person responsible for answering queries on the specimens.
6. **CONFIRMATION:** For every shipment, the primary investigator leading the project must confirm that all samples contained within the shipment have been collected with local permissions and under local ethical guidelines. Please write “YES” in this column to confirm that these standards have been met and that the necessary regulatory compliance documents have been obtained and are available to you, only once is necessary, and in the same row as the primary investigator’s details. We are not able to accept samples that do not meet these standards. These may include landowner permission, veterinary pathogen sampling permissions, and/or Nagoya compliance. This is an important check that ensures that permissions were granted to collect and transfer the specimen for this research purpose. The sample provider should ensure this documentation is obtained, and that copies of the relevant paperwork are shared with Sanger where necessary and as stipulated, for example, by regulations/approvals or licensing authorities.

# 6. Tab2: Column by column instructions for the Metadata Entry tab

* 1. **SERIES:** This field is simply a series of numbers that should reflect the total number of wells submitted, e.g., if 10 plates are submitted, each with 95 samples (and one blank sample well), then the series will begin at 1 and end at 960.
  2. **CATCH\_LOT:** One 50 mL falcon tube containing 100% ethanol is used per trap bottle for each Malaise trap collection event (or other en masse trap collection event). That is, for a bidirectional Malaise trap, two 50 mL tubes will be used, one for each trap bottle during the catch period. Each 50 mL tube has an alphanumeric label barcode set up as C###[A-Z], e.g. C001A. Any leftover or excess insects remaining after plating should be returned to the same 50 mL falcon tube along with any remaining ethanol from the trap.
  3. **RACK\_OR\_PLATE\_ID:** Each plate must be labelled with a sticker that has a human readable label and a scannable barcode. The human readable label will contain your Partner Code, which we will assign to you, and a numeric plate identifier (e.g. CAMP-001 would represent the first plate from the CAMPUS site). We will provide you with the stickers to label your own plates or with labelled plates if you require these materials from us.
  4. **TUBE\_OR\_WELL\_ID:** This field should have a record for each well in a plate. Adding the WELL\_ID for samples in plates can be expedited by copy/pasting from the appropriate column in TAB 3, the Well IDs tab. If plate level only metadata is being entered then please copy and paste “PLATE\_ONLY” from Tab3 into this field and use only one row to capture the metadata for the whole plate. Use this only if a plate represents all identical metadata (i.e. one catch lot, all specimens are WHOLE\_ORGANISM, there are no taxonomic identifications, no blank samples other than H12, and all specimens are from one catch lot). Otherwise please follow previous guidance and submit 96 rows of metadata.
  5. **ORGANISM\_PART:** A description of the exact tissue(s) in the tube or well. This field has a controlled vocabulary: use the drop-down menu. For the empty well required on each plate that should contain only the preservative solution (or for any other wells that do not contain specimens), select “*NOT\_APPLICABLE*” (H12). Entering “*NOT\_APPLICABLE*” for empty wells is critical as it informs us that there is no sample expected in the well. Refer to *Section 4* for guidance on how to plate larger insects.
     1. If multiple parts must be submitted within a well, combine tissues by entering multiple terms from the ontology using the | (vertical pipe) symbol (e.g. for head + abdomen of an insect enter “HEAD | ABDOMEN”). When using multiple body parts, there will be a data validation error that arises, but these can be ignored as long as the spelling and capitalization of the terms are identical to the provided list. If some damage has been done but the specimen is largely complete (e.g., missing some legs or wings), then select “*WHOLE\_ORGANISM*”.
  6. **PRESERVATIVE\_SOLUTION:** Suspension liquid used to preserve each sample in the 96 well plate, which should typically be “100%\_ETHANOL”. This field has a controlled vocabulary: select the correct option from the drop down menu for the first entry, then in order to fill multiple samples, click on the cell containing the data entry and then click on the little blue square that will appear at the lower right corner of the cell (in Google Sheets) and drag the square down to fill any remaining cells with the same information. If a different percentage of ethanol has been used, select “V%\_ETHANOL” and add the percentage (e.g. 85% ethanol) to the field PRESERVATION\_APPROACH. If none of the drop-down terms are suitable, select “OTHER” and add information to PRESERVATION\_APPROACH.
  7. **BOTTLE\_DIRECTION:** When the catch comes from a Malaise trap, please put the direction the bottle end of the trap was oriented towards using the one (N,S,W,E), two (NE,SE,NW,SW), or three letter codes as below. If using a two-bottle trap then label each bottle with the corresponding compass point with the ethanol proof paint marker provided, so you know which bottle has each catch when you remove them for plating. Leave this field empty if the specimen is not from a Malaise trap.



* 1. **DATE\_OF\_COLLECTION:** The date of the sample collection, with year, month and day specified in this order using hyphens (YYYY-MM-DD). It is important to provide the complete date, but if only the year or the year and month are known, then enter YYYY or YYYY-MM, respectively. If the collection spans multiple days (as Malaise catches set for 24 hours will), we request that the **end** date is entered here rather than the start date. If your collection ends precisely at midnight, we consider midnight to be the first minute of the day (00:00 as opposed to 24:00) and thus the end date should reflect this.
  2. **COUNTRY\_OF\_COLLECTION:** This must use the accepted country name, which can be looked up here <https://www.insdc.org/country>.
  3. **COLLECTION\_LOCATION:** Where possible, this should use your country’s administrative districts to as fine of a resolution as you can provide ranging from least to most specific and separated by | character, e.g. “*England | East of England | Cambridgeshire | Hinxton | Wellcome Genome Campus | East Pond*”. It is important to give the name of the specific collection site whenever possible. If the specimen is from a laboratory colony, give the current location (e.g. Imperial College London) rather than information on the original collection, but indicate original collection location in OTHER\_INFORMATION.
  4. **DECIMAL\_LATITUDE:** In decimal degrees, between -90 and 90. We advise that locations are specified to at least 4 decimal places (<https://en.wikipedia.org/wiki/Decimal_degrees>) giving 11 m resolution to the location from which the Malaise catch was collected. If GPS coordinates were not taken at the time of collection, this is easy to retrospectively collect by dropping a pin in google maps or grid reference finder for UK sites (<https://gridreferencefinder.com/>) and clicking on the pin to reveal coordinates.
  5. **DECIMAL\_LONGITUDE:** guidance same as latitude, except that range is from -180 to 180.
  6. **WHAT\_3\_WORDS:** Optional field to include information to geolocate the sample area to a 3 m square, mapped by what3words.com (see guidance at <https://what3words.com/how-to-use-the-what3words-app/>), this can also be found using site <https://gridreferencefinder.com/>. This can be used to give more precise resolution to where the Malaise trap was set up or to distinguish between nearby sampling locations or mark your Malaise trap location more precisely in cases where the trap must be disassembled and reassembled each month. Please include the three backslashes as part of the entry, e.g., ///protected.cheetahs.slippery.
  7. **TIME\_OF\_COLLECTION:** Time of day at which the collection **ended** in local time in 24-hour clock format, with hours and minutes separated by colon e.g. 13:35, 04:53, etc. We do not accept 24:00 for midnight and you must use 00:00 (first minute of the day) if your collection ends precisely at midnight. This should be in local time.
  8. **DURATION\_OF\_COLLECTION:** This field captures the duration of the catch (how long the trap was set up and operational) using ISO 8601 standards. The format is complicated and described below in detail – briefly, if you have set a trap for a 24 hour period, enter PT24H. Round up to the nearest hour - we do not require minutes or seconds to be recorded in this field. If you have set a trap for 20 hours and 30 minutes, enter PT21H, etc.

Duration format in detail in case of activities that require more specific time records: The format is P[n]Y[n]M[n]DT[n]H[n]M[n]S. In this representation, the [n] is replaced by the value for each of the date and time elements that follow the [n]. The capital letters *P*, *Y*, *M*, *W*, *D*, *T*, *H*, *M*, and *S* are designators for each of the date and time elements and are not replaced.

*P* is the duration designator (for *period*) and is always placed at the start of the duration representation.

* + *Y* is the year designator that follows the value for the number of years.
  + *M* is the month designator that follows the value for the number of months.
  + *W* is the week designator that follows the value for the number of weeks.
  + *D* is the day designator that follows the value for the number of days.

*T* is the time designator that precedes the time components of the representation.

* + *H* is the hour designator that follows the value for the number of hours.
  + *M* is the minute designator that follows the value for the number of minutes.
  + *S* is the second designator that follows the value for the number of seconds.
  1. **COLLECTION\_METHOD:** A description of the collection method used. This field has a controlled vocabulary; pick the best fitting term from the drop down menu. Any details about the collection method can be recorded in the column DESCRIPTION\_OF\_COLLECTION\_METHOD. If none of the available terms describe your collection method, select \*\*OTHER\*\* and describe the method in the column DESCRIPTION\_OF\_COLLECTION\_METHOD or request an additional term to be added.
  2. **DATE\_OF\_PLATING:** Enter the date when the Malaise trap samples (or other *en masse* traps), were plated. This can be the same day as DATE\_OF\_COLLECTION.
  3. **PREDICTED\_ORDER\_OR\_GROUP:** The taxonomic Order or higher classification into which the Family and Genus belongs. If you are not confident in the identification then indicate this as “Y” in the SPECIMEN\_IDENTITY\_RISK field. This should correspond to the taxonomy as represented in the NCBI Taxonomy Database.
  4. **PREDICTED\_FAMILY:** The taxonomic Family (including Superfamily, Subfamily, Tribe or Subtribe) into which the Genus is placed. This should correspond to the taxonomy as represented in the NCBI Taxonomy Database.
  5. **PREDICTED\_GENUS:** The taxonomic Genus (including Subgenus) to which the Species belongs. This should correspond to the taxonomy as represented in the NCBI Taxonomy Database, and with the generic component of the scientific name given below.
  6. **PREDICTED\_SCIENTIFIC\_NAME:** The latin binomial/combined genus and species name with a space in between. If the specimen has not been identified to species level, leave the field empty. All samples will then be registered as “*unidentified*” in the validation process. If the well does not contain a sample (e.g. H12), again this field can be left empty as long as ORGANISM\_PART has been registered as “NOT\_APPLICABLE”.
  7. **SPECIMEN\_IDENTITY\_RISK:** Y/N field to indicate if there is any risk that the specimen is part of a species complex or group where it can be difficult to be certain of species identity and/or species boundaries. Enter “N” (i.e. no risk) if there is high confidence that the specimen has been identified to the correct species level as recorded in PREDICTED\_SCIENTIFIC\_NAME. If you can identify the specimen to a species group or complex, but perhaps not all the way to species level, enter “Y” (i.e., yes there is a risk the specimen is not accurately identified to species level resolution).
  8. **LIFESTAGE:** The lifestage of the specimen from which the sample was derived at the time it was preserved (if this is different than the lifestage when collected e.g. if larvae were collected in the field and reared to adults in the lab, the entry here should be adult). This field has a controlled vocabulary, please use the drop-down menu. If you would like to provide more detailed lifestage information than is available in the drop-down menu, add this to “OTHER\_INFORMATION” (e.g., L3 stage larva).
  9. **SEX**: The sex of the specimen from which the sample was derived. This field has a controlled vocabulary: use the drop-down menu. If the sex of the organism is not known, leave this field empty.
  10. **SORTING\_SOLUTION\_USED:** Indicate if you used the phosphate buffered saline (PBS) solution to morphosort your specimens before plating. This field has a controlled vocabulary, please use the drop-down menu.
  11. **CATCH\_BOTTLE\_TEMPERATURE\_STORAGE:** Indicate the temperature at which the Catch Bottle was stored for the majority of the time before specimens were plated. This field has a controlled vocabulary, please use the drop-down menu.
  12. **PLATE\_TEMPERATURE\_STORAGE:** Indicate the temperature at which the specimen plates were stored for the majority of the time before plates were shipped to Sanger. This field has a controlled vocabulary, please use the drop-down menu.
  13. **MORPHOSPECIES\_DESCRIPTION:** This is a free text field for you to use to write in any descriptive words for morphospecies that you come across often in your catches, however you are unable to identify them, e.g. “brown fly with yellow stripe on abdomen”. This information can be useful for you to match up species data when you receive your biodiversity report and useful to us when reviewing the sequence data output.
  14. **DESCRIPTION\_OF\_COLLECTION\_METHOD:** A detailed as possible description of the sample collection methods, e.g. “*caught with fibre net within densely wooded area, and immediately placed into the collection container”* or “*manually collected whilst feeding on a cow inside a shed and immediately placed into the collection container”*. If the specimen was collected as larvae and reared to adult, mention this here.
  15. **HABITAT**: Free text field to include any comments about the location, habitat or substrate, *e.g. damp mossy ground in moderate shade* or *indoors in air conditioned office space.* This is a work in progress and can use ENVO terms, NVC, or EUNIS.
  16. **PRESERVATION\_APPROACH:** This is free text but should summarise any extra information regarding the temperature and storage time you may have from the point of collection to the point of sample being placed in the well, for example, Catch bottle kept in fridge for 2 weeks, then moved to room temperature (RT) for 2 days before specimens were plated out and then plates remained at RT until shipping.
  17. **COLLECTOR\_SAMPLE\_ID:** This is the unique name assigned to the sample by the COLLECTOR. This is a free text field, but please do not use spaces or special characters, other than hyphens and underscores (“-” and “\_”) i.e do not use #, !, ^, \*, etc.
  18. **VOUCHER\_ID:** Where the entire specimen is not submitted, the remaining specimen can be pinned or stored in preservative. This remaining specimen should have an associated and unique VOUCHER\_ID that should be determined by the Partner such that sequencing data can be linked back to the Voucher Specimen at a later date. When the whole insect is submitted for sequencing, this field should be left empty.
  19. **ELEVATION:** Altitude above sea level, supplied in metres. Do not supply the unit, e.g. use 200 for 200 m above sea level, 100-200 for 100-200 m range above sea level, etc. Please supply elevation of water surface for inland water bodies. Leave this field empty if the elevation was not recorded.
  20. **OTHER\_INFORMATION:** Free text field for further relevant information not captured by the other fields. Scientific names of other species expected to be present in this specific sample, taking consideration of the ORGANISM\_PART. Occasionally, the relationships between submitted samples may be known and important. This can be reflected in the COLLECTOR\_SAMPLE\_ID and described here (e.g. AB43\_F1 and AB43\_Mother). If there is nothing else to add here, this field should be left empty.
  21. **MISC\_METADATA:** Please discuss this by emailing [bioscan.info@sanger.ac.uk](mailto:bioscan.info@sanger.ac.uk) before filling in the manifest if additional data for your samples exists but is not captured by any existing fields. In such a case, we should make every effort to collect and standardise additional metadata. This column can be copy/pasted to create as many additional columns as required to capture any additional metadata of interest, for example, this might include temperature, humidity, windspeed, baited traps. Please change the field from MISC\_METADATA to the proposed name of the new field, and standardise your entries using ALL CAPs and no spaces, e.g. a new field might be “BAITED\_TRAP”and terms might be “CARRION”, “OVERRIPE\_FRUIT”.
  22. **IDENTIFIED\_BY:** Enter the name of the person or people who identified the specimen, only if identified down to species level. Use ALL CAPs, and separate names with | (vertical pipe symbol), e.g., “CAROLUS LINNAEUS | JEAN-BAPTISTE LAMARCK”. If the specimen was not identified to species level, then put “NOT\_APPLICABLE” here.

If a contributor has identified a specimen to species name, the columns that capture the information on identifiers can be found at the end of the spreadsheet in TAB 2, we will only use this information to contact the identifier if there has been a discrepancy with the scientific name (e.g. spelling).

* + 1. We note that storage of names with affiliations in a database brings the system under the aegis of the GPDR regulations, and we must ask partners to agree to their data being stored and to those data being propagated to secondary databases (including ENA and the final collections of record).
  1. **IDENTIFIER\_AFFILIATION:** Free text field to supply the university, institution, or society that is responsible for the collected specimen. This is typically the society or institution of the person(s) specified in the IDENTIFIED\_BY field. If multiple people are specified in IDENTIFIED\_BY, ensure that their institutional affiliations are also separated by a vertical pipe symbol. Position in the list of affiliations should match the person in the same position in the list of names (e.g. “Person A | Person X | Person C” will have their affiliations as: “Institute A | Institute X | Institute C”. If multiple people are listed but all from the same affiliation, no need to repeat the affiliation. If the specimen was not identified to species level, then put “NOT\_APPLICABLE” here.

# 7. The Post-validation Process

All fields listed above require metadata to be filled in by specimen providers. Once validation has been completed, additional fields are created as described below and a copy of this final manifest will be returned to you.

**TAXON\_ID:** This is recorded in all cases as “unidentified” 32644 or “blank sample” 2582415.

**SCIENTIFIC\_NAME:** This is recorded in all cases as “unidentified” or “blank sample” depending on whether there was a specimen present in the well or not. Empty wells will then be registered as “*blank sample*” to differentiate from “*unidentified*” in the validation process. This is a formality to ensure that we are able to update to the correct name after sequencing is complete.

**SPECIMEN\_ID:** This is a unique identifier given to each specimen through combining RACK\_OR\_PLATE\_ID + ‘\_’ + TUBE\_OR\_WELL\_ID (e.g., AYDI\_001\_A1)

**PREDICTED\_TAXON\_ID:** This is the NCBI Taxonomy identifier assigned to the specimen based on the PREDICTED\_SCIENTIFIC\_NAME – if no scientific name is provided, this will be recorded as 32644. If a species name is provided, the correct TAXON\_ID will be added to this field.

**CATCH\_SOLUTION:** This is the suspension liquid used to preserve the sample in the Malaise catch bottles, this is provided by the BIOSCAN team and added for internal metadata tracking.

**GAL, SYMBIONT, REGULATORY\_COMPLIANCE, and HAZARD\_GROUP** may all also be added upon completion of validation because they are required by our internal sample tracking system.

Document History

| ***Version*** | ***Date*** | ***Changes*** | ***Contributors*** |
| --- | --- | --- | --- |
| **1.0** | May 2021 | First version of the BIOSCAN manifest | Lyndall Pereira, Vickie Brookes, Mara Lawniczak, Marilou Boddé, Petra Korlevic, Katie Woodcock |
| **2.0** | June 2022 | Requirements for missing data terms removed.  Allow partial plates for final catch before shipment.  Tabs moved around.  Epicollect not to be used, Google folder - each partner code with own link.  Allow plate level entry into the same manifest - called PLATE\_ONLY under TUBE\_OR\_WELL\_ID column.  Added CONFIRMATION to Contributors tab.  WHOLE\_ORGANISM\_DAMAGED and BLANK\_SAMPLE removed from ORGANISM\_PART, BLANK\_SAMPLE to be replaced with NOT\_APPLICABLE to show empty wells.  Post modification fields added at the end | Lyndall Pereira, Mara Lawniczak, Petra Korlevic, Alex Makunin |