# SYSTEMATIC SURVEY METADATA TEMPLATE INSTRUCTIONS

## OVERVIEW

Use this template to record systematic survey metadata; that is metadata relating to a Systematic Survey dataset.

This Systematic Survey Metadata template must be used in combination with the Systematic Survey Occurrence template and, in some cases, the Systematic Survey Site template.

Templates have been provided to facilitate integration of your data into the Biodiversity Data Repository (BDR) database. Not all types of data have been catered for in the available templates at this stage; therefore, if you are unable to find a suitable template, please contact bdr-support@gaiaresources.com.au to make us aware of your data needs.

### NEED TO KNOW:

For data validation, you will need your data file to:

- be the correct **file format**,
- have **matching template fields** to the template downloaded (do not remove, or change the order of fields),
- additional fields may be added **after the templated fields** (noting that the data type is not assumed and values will be encoded as strings),
- only one row of metadata should be included and **only the first row of metadata will be accepted** (this symbolises one Survey per dataset submission).
- have values in **mandatory fields** (see Table 1), and
- comply with data **value constraints,** for example the geographic coordinates are consistent with a geodeticDatum type of the five available options.

## FILE FORMAT

- The systematic survey metadata template is a UTF-8 encoded csv (not Microsoft Excel Spreadsheets). Be sure to save this file with your data as a .csv (UTF-8) as follows, otherwise it will not pass the in-browser csv validation step upon upload.
  [MS Excel: Save As > More options > Tools > Web options > Save this document as > Unicode (UTF-8)]
- **Do not include empty rows**.

### FILE SIZE

MS Excel imposes a limit of 1,048,576 rows on a spreadsheet, limiting a CSV file to the header row followed by 1,048,575 occurrences. Furthermore, MS Excel has a 32,767 character limit on individual cells in a spreadsheet. These limits may be overcome by using or editing CSV files with other software.

# TEMPLATE FIELDS

The template file contains the field names in the top row that form part of the core Survey data model. Table 1 will assist you in transferring your data to the template with the following information:

- **Field name** in the template (and an external link to the Darwin Core standard for that field where available);
- **Description** of the field;
- **Required** i.e. whether the field is **mandatory or optional**;
- **Datatype format** required for the data values for example text (string), number (integer, float), or date; and
- **Example/s** of an entry for that field.
- **Vocabulary links** within this document (for example pick list values) where relevant. The fields that have suggested values options for the fields in Table 1 are listed in Table 2 in alphabetical order of field name.

# ADDITIONAL FIELDS

Data that do not match the existing template fields may be added as additional columns in the CSV files after the templated fields.
E.g., sampleSizeUnit, sampleSizeValue.

Table 1: Systematic Survey Metadata template fields with descriptions, conditions, datatype format, and examples.

| Field name | Description | Mandatory / Optional | Datatype Format | Examples / Vocabulary |
|---|---|---|---|---|
| surveyID | The identifier for the survey. Important if there is more than one survey in the project. | Optional | String | COL1 |
| surveyName | Brief title for the survey. | Mandatory | String | Disentangling the effects of farmland use, habitat edges, and vegetation structure on ground beetle morphological traits - Summer |
| surveyPurpose | A description of the survey objective. | Optional | String | Summer sampling for peak insect diversity. |
| surveyType | Description of type of survey conducted. | Optional | String | Wet pitfall trapping |
| surveyStart | The date data collection commenced. | Mandatory | Timestamp | 23/09/2020 |
| surveyEnd | The date data collection was completed. | Optional | Timestamp | 23/09/2020 |
| targetTaxonomicScope | The range of biological taxa covered by the survey. Multiple terms are allowed, separated by a vertical bar aka pipe \| | Optional | List | Coleoptera \| Formicidae |
| targetHabitatScope | The habitats targeted for sampling during the survey. Multiple terms are allowed, separated by a vertical bar aka pipe \| | Optional | List | Woodland |
| spatialCoverageWKT | Well Known Text (WKT) expression of the geographic coordinates that describe the survey's spatial extent. | Optional | WKT | POLYGON ((146.363 -33.826, 148.499 -33.826, 148.499 -34.411, 146.363 -33.826)) (WKT notes) |

| Field name | Description | Mandatory / Optional | Datatype Format | Examples / Vocabulary |
|---|---|---|---|---|
| geodeticDatum | The geodetic datum upon which the geographic coordinates in the Spatial coverage (WKT) are based. | Conditionally mandatory with spatialCoverageWKT | String | GDA2020 (Vocabulary link) |
| surveyOrgs | Name of organisation or individual for whom Survey is being conducted. Multiple terms are allowed, separated by a vertical bar aka pipe \| | Optional | List | NSW Department of Planning, Industry and Environment \| CSIRO |
| samplingPerformedBy | Similar to eco:samplingPerformedBy. Individual/s that carried out the survey. These can be represented by identifiers. Multiple values allare allowed, separated by a vertical bar aka pipe \| | Optional | List | https://orcid.org/0000-0003-1092-9705 \| https://orcid.org/0009-0004-5278-4238 |
| surveyMethodCitation | A citation or reference to the survey methods used. | Optional | List | Ng, K., Barton, P.S., Blanchard, W. et al. Disentangling the effects of farmland use, habitat edges, and vegetation structure on ground beetle morphological traits. Oecologia 188, 645–657 (2018) \| https://doi.org/10.1007/s00442-018-4180-9 |

| Field name | Description | Mandatory / Optional | Datatype Format | Examples / Vocabulary |
|---|---|---|---|---|
| surveyMethodDescription | Free text description of the survey method used. | Optional | String | Our experimental design consisted of four 400 m transects running from inside each woodland patch out into four adjoining farmland uses (crop, rested, woody debris application, revegetation plantings). To quantify potential edge effects on beetle species traits, we sampled beetles at five locations along each transect: 200 and 20 m inside woodlands, 200 and 20 m inside farmlands, and at the woodland–farmland edge (0 m). Each sampling location comprised a pair of wet invertebrate pitfall traps. separated by a drift fence (60 cm long x 10 cm high) to help direct arthropods into traps. We opened a total of 220 pairs of traps for 14 days during spring (Oct–Nov 2014), and repeated sampling during summer (January–February 2015). Beetle samples from each pitfall trap pair, and across the two time periods, were pooled to provide one sample per sampling location. |
| surveyMethodURL | A DOI or link to the reference about the survey method, if available. | Optional | List | https://biocollect.ala.org.au/document/download/2022-01/202201%20CBR%20Flora%20and%20Vegetation%20report_draftv1.pdf \| https://doi.org/10.1002/9781118945568.ch11 |

| Field name | Description | Mandatory / Optional | Datatype Format | Examples / Vocabulary |
|---|---|---|---|---|
| samplingEffortValue | Similar to eco:samplingEffortValue. The total sampling effort value. A samplingEffortValue must have a corresponding samplingEffortUnit | Optional | String | 20 x 12 |
| samplingEffortUnit | Similar to eco:samplingEffortUnit. The units associated with samplingEffortValue. | Conditionally mandatory with samplingEffortValue | String | trapDays |
| keywords | Terms, phrases or descriptors that highlight the key attributes of the study. Multiple terms are allowed, separated by a vertical bar aka pipe \| | Optional | List | ground beetle \| habitat \| morphology \| traits \| farmland \| woodland \| remnant vegetation \| split-plot study |

# APPENDICES

## APPENDIX-I: VOCABULARY LIST

Table 2 describes preferred labels for geodeticDatum.

Note: The geodeticDatum value must come from one of the Preferred labels or Alternate Labels in this table.

Table 2: Suggested values for geodeticDatum controlled vocabulary fields in the template.

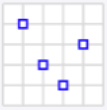
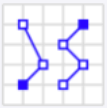
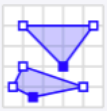
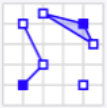
| Template field name | Preferred label | Definition | Alternate label |
|---|---|---|---|
| geodeticDatum | AGD84 | Australian Geodetic Datum 1984 | EPSG:4203 |
| | AGD66 | Australian Geodetic Datum 1966 | EPSG:4202 |
| | GDA2020 | Geocentric Datum of Australia 2020 | EPSG:7844 |
| | GDA94 | Geocentric Datum of Australia 1994 | EPSG:4283 |
| | WGS84 | World Geodetic System 1984, used in GPS | EPSG:4326 |

# APPENDIX-II: Well Known Text (WKT)

For general information on how WKT coordinate reference data is formatted here. The length of a WKT string or of its components is not prescribed. However MS Excel has a 32,767 (32K) character limit on individual cells in a spreadsheet.

It is possible to edit CSV files outside of Excel in order to include more than 32K characters.

**Multipart geometries (2D)**

| Type | Examples | |
|---|---|---|
| MultiPoint |  | `MULTIPOINT ((10 40), (40 30), (20 20), (30 10))` |
| | | `MULTIPOINT (10 40, 40 30, 20 20, 30 10)` |
| MultiLineString |  | `MULTILINESTRING ((10 10, 20 20, 10 40), (40 40, 30 30, 40 20, 30 10))` |
| MultiPolygon |  | `MULTIPOLYGON (((30 20, 45 40, 10 40, 30 20)), ((15 5, 40 10, 10 20, 5 10, 15 5)))` |
| |  | `MULTIPOLYGON (((40 40, 20 45, 45 30, 40 40)), ((20 35, 10 30, 10 10, 30 5, 45 20, 20 35), (30 20, 20 15, 20 25, 30 20)))` |
| GeometryCollection |  | `GEOMETRYCOLLECTION (POINT (40 10), LINESTRING (10 10, 20 20, 10 40), POLYGON ((40 40, 20 45, 45 30, 40 40)))` |

# APPENDIX-III: **Timestamp**

Following date and date-time formats are acceptable within the timestamp:

| | |
|---|---|
| **xsd:dateTimeStamp with timezone** | yyyy-mm-ddThh:mm:ss.sTZD (eg 1997-07-16T19:20:30.45+01:00) OR<br>yyyy-mm-ddThh:mm:ssTZD (eg 1997-07-16T19:20:30+01:00) OR<br>yyyy-mm-ddThh:mmTZD (eg 1997-07-16T19:20+01:00) |
| **xsd:dateTime** | yyyy-mm-ddThh:mm:ss.s (eg 1997-07-16T19:20:30.45) OR<br>yyyy-mm-ddThh:mm:ss (eg 1997-07-16T19:20:30) OR<br>yyyy-mm-ddThh:mm (eg 1997-07-16T19:20) |
| **xsd:Date** | dd/mm/yyyy OR<br>d/m/yyyy OR<br>yyyy-mm-dd OR<br>yyyy-m-d |
| **xsd:gYearMonth** | mm/yyyy OR<br>m/yyyy OR<br>yyyy-mm |
| **xsd:gYear** | yyyy |

Where:

yyyy = four-digit year
mm = two-digit month (01=January, etc.)
dd = two-digit day of month (01 through 31)
hh = two digits of hour (00 through 23) (am/pm NOT allowed)
mm = two digits of minute (00 through 59)
ss = two digits of second (00 through 59)
s = one or more digits representing a decimal fraction of a second
TZD = time zone designator (Z or +hh:mm or -hh:mm)

# APPENDIX-IV: UTF-8

UTF-8 encoding is considered a best practice for handling character encoding, especially in the context of web development, data exchange, and modern software systems. UTF-8 (Unicode Transformation Format, 8-bit) is a variable-width character encoding capable of encoding all possible characters (code points) in Unicode.
Here are some reasons why UTF-8 is recommended:**Universal Character Support:** UTF-8 can represent almost all characters from all writing systems in use today. This includes characters from various languages, mathematical symbols, and other special characters.

- **Backward Compatibility:** UTF-8 is backward compatible with ASCII (American Standard Code for Information Interchange). The first 128 characters in UTF-8 are identical to ASCII, making it easy to work with systems that use ASCII.
- **Efficiency:** UTF-8 is space-efficient for Latin-script characters (common in English and many other languages). It uses one byte for ASCII characters and up to four bytes for other characters. This variable-length encoding minimises storage and bandwidth requirements.
- **Web Standards:** UTF-8 is the dominant character encoding for web content. It is widely supported by browsers, servers, and web-related technologies.
- **Globalisation:** As software applications become more globalised, supporting a wide range of languages and scripts becomes crucial. UTF-8 is well-suited for internationalisation and multilingual support.
- **Compatibility with Modern Systems:** UTF-8 is the default encoding for many programming languages, databases, and operating systems. Choosing UTF-8 helps ensure compatibility across different platforms and technologies.

When working with text data, it's generally a good idea to use UTF-8 encoding to avoid issues related to character representation and ensure that your software can handle a diverse set of characters and languages.

For assistance, please contact: bdr-support@gaiaresources.com.au