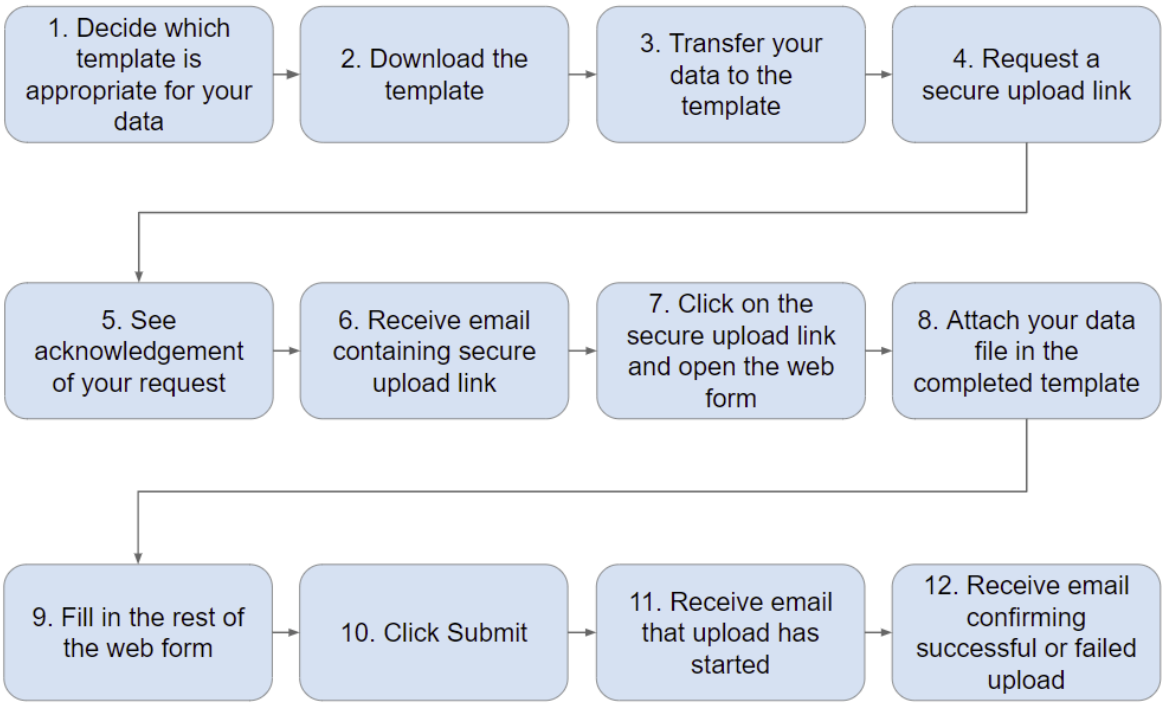
Instructions - Submitting CSV Data to the Biodiversity Data Repository

Overview

A set of templates is provided for you to map your terrestrial flora and/or fauna data to in order to submit it to the DAWE Biodiversity Data Repository (BDR). These instructions for the **Species Occurrence (CSV)** template explain how the template works, what the fields mean and the steps you take to prepare your biodiversity data for successful upload.

Steps

The process of data submission to the BDR is outlined below. The first three steps involve the downloading of a template and populating it with the records you wish to submit. The steps that follow relate to the actual submission process on the BDR website.



Step 1: Decide which template is appropriate for your data

Depending on how you manage your biodiversity data, you may prefer to use either the CSV or Geopackage template provided. The CSV template is best suited to records in tabular form, while the Geopackage update suits records in spatial formats.

For example, if you work primarily with spreadsheets and tables, then the CSV template provides a common format that can be used across a range of spreadsheet software

applications. If you prefer to manage your data in a Geographic Information System (GIS) then the Geopackage template provides an open-source format that can be used across common GIS software applications.

Both formats contain the same required fields that are based on the Darwin Core standard.

Note that in future we intend to provide a greater number of templates to suit other common types of biodiversity information; but you will be able to access them all from this same site.

Step 2: Download the template

The next step is to download the template itself so that you can work with it. Biodiversity data templates are provided from this website:
https://staging-bdr.gaiaresources.com.au/#/template/download

Save the template on your computer.

Step 3: Transfer your data to the template

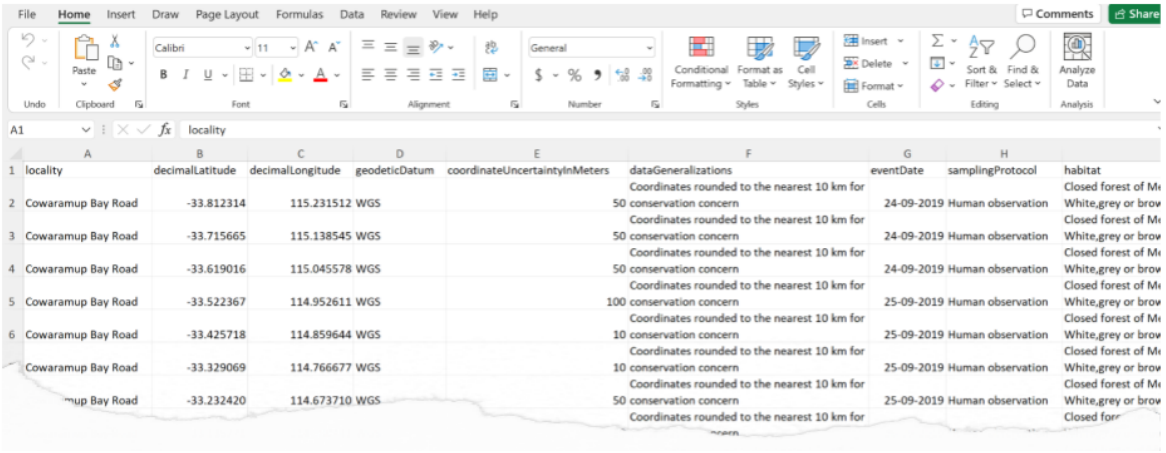
Now comes the hard work of transferring your biodiversity data to the template. The CSV template can easily be opened in Microsoft Excel, Libre Office or another spreadsheet software. From there, each occurrence record should be a single row in your template. To help you in that task, the table below provides a number of resources for each field:

- The name of the field (and an external link to the Darwin Core standard for that field)
- The description of the field
- The required data format as text (e.g. string), number (eg. integer, float) or date
- Whether the field is mandatory - where all records require an entry - or optional
- An example of an entry for that field
- Where applicable, we also provide a link to the controlled vocabularies (e.g. pick list values) for that field.

In your own organisation, you may have a mix of historical data and data collection processes, resulting in different field names and formats. If the available templates do not cover your data requirements adequately, we can work with you to create a customised template.

Template Fields

Let's have a look at the fields and what they mean - below is an example, and Table 1 provides a full description of each field.



	locality	decimalLatitude	decimalLongitude	geodeticDatum	coordinateUncertaintyInMeters	dataGeneralizations	eventDate	samplingProtocol	habitat
1						Coordinates rounded to the nearest 10 km for			Closed forest of M
2	Cowaramup Bay Road	-33.812314	115.231512	WGS	50	conservation concern	24-09-2019	Human observation	White,grey or brov
3	Cowaramup Bay Road	-33.715665	115.138545	WGS	50	Coordinates rounded to the nearest 10 km for conservation concern	24-09-2019	Human observation	Closed forest of M White,grey or brov
4	Cowaramup Bay Road	-33.619016	115.045578	WGS	50	Coordinates rounded to the nearest 10 km for conservation concern	24-09-2019	Human observation	Closed forest of M White,grey or brov
5	Cowaramup Bay Road	-33.522367	114.952611	WGS	100	Coordinates rounded to the nearest 10 km for conservation concern	25-09-2019	Human observation	Closed forest of M White,grey or brov
6	Cowaramup Bay Road	-33.425718	114.859644	WGS	10	Coordinates rounded to the nearest 10 km for conservation concern	25-09-2019	Human observation	Closed forest of M White,grey or brov
	Cowaramup Bay Road	-33.329069	114.766677	WGS	10	Coordinates rounded to the nearest 10 km for conservation concern	25-09-2019	Human observation	Closed forest of M White,grey or brov
	up Bay Road	-33.232420	114.673710	WGS	50	Coordinates rounded to the nearest 10 km for conservation concern	25-09-2019	Human observation	Closed forest of M White,grey or brov
						Coordinates rounded to the nearest 10 km for			Closed for

There are several mandatory fields that will require a value in every entry, and many additional optional ones. Mandatory fields will need to be filled out for the submission to be accepted, at times, using the required controlled value as indicated.

There are some fields that have suggested values (available at https://etc) that may aid the user in how to populate those fields; however, the value options are not enforced.

Some fields have controlled vocabularies, which means that your records will need to conform to an entry from a list of accepted values, to be accepted by the BDR system. If a field is optional, but has a controlled vocabulary, then a blank (e.g. Null) value will still be accepted.

A few of the fields also have range constraints that are put in place to ensure their values are sensible. For example, the location of a record is required as a latitude and longitude in decimal degrees format, and there is a valid range of these values that locate the record within either the Australian continent or Australia's island territories. These ranges are checked by the system upon submission.

Note the BDR system is designed to only accept terrestrial biodiversity data at this time; while marine biodiversity data is of interest, in the system's current implementation, marine data may be flagged and ultimately not included in the BDR system.

Table 1: Occurrence field descriptions

Field name	Description	Mandatory / Optional	Format	Example / Vocabulary
locality	The specific description of the place.	Optional	String	Cowaramup Bay Road
decimalLatitude	The geographic latitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are	Mandatory	Float	-33.812314

	north of the Equator, negative values are south of it. Legal values lie between -90 and 90, inclusive.			
decimalLongitude	The geographic longitude (in decimal degrees, using the spatial reference system given in geodeticDatum) of the geographic centre of a Location. Positive values are east of the Greenwich Meridian, negative values are west of it. Legal values lie between -180 and 180, inclusive.	Mandatory	Float	115.231512
geodeticDatum	The EPSG code for the ellipsoid, geodetic datum, or spatial reference system (SRS) upon which the geographic coordinates given in decimalLatitude and decimalLongitude as based.	Mandatory	String	WGS (Vocabulary link)
coordinateUncertaintyMeters	The horizontal distance (in metres) from the given decimalLatitude and decimalLongitude describing the smallest circle containing the whole of the Location. Leave the value empty if the uncertainty is unknown, cannot be estimated, or is not applicable (because there are no coordinates). Zero is not a valid value for this term.	Optional	Integer	50
dataGeneralisations	Actions taken to make the shared data less specific or complete than in its original form. Suggests that alternative data of higher quality may be available on request.	Optional	String	Coordinates rounded to the nearest 10 km for conservation concern
eventDate	The date-time (in ISO 8601 format) or interval during which an Event occurred. For occurrences, this is the date-time when the event was recorded. Not suitable for a time in a geological	Mandatory	Date (ISO)	2019-09-24

	context.			
samplingProtocol	Recommended best practice is describe an Event with no more than one sampling protocol. In the case of a summary Event in which a specific protocol can not be attributed to specific Occurrences, the recommended best practice is to repeat the property for each IRI that denotes a different sampling protocol that applies to the Occurrence.	Mandatory	String	Human observation (Vocabulary link)
habitat	A category or description of the habitat in which the Event occurred.	Optional	String	Closed forest of Melaleuca lanceolata. White,grey or brown sand, sandy loam.
recordedBy	A person, group, or organisation responsible for recording the original Occurrence.	Optional	String	J. Doe
basisOfRecord	The specific nature of the data record.	Optional	String	Human observation
occurrenceID	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.	Optional	String	S09-20190924-02
individualCount	The number of individuals present at the time of the Occurrence.	Optional	Integer	1
lifeStage	The age class or life stage of the Organism(s) at the time the Occurrence was	Optional	String	Mature (Vocabulary

	recorded.			link)
sex	The sex of the biological individual(s) represented in the Occurrence.	Optional	String	Unspecified (Vocabulary link)
reproductiveCondition	The reproductive condition of the biological individual(s) represented in the Occurrence.	Optional	String	flowering (Vocabulary link)
establishmentMeans	Statement about whether an organism or organisms have been introduced to a given place and time through the direct or indirect activity of modern humans.	Optional	String	native
organismRemarks	Comments or notes about the Organism instance.	Optional	String	good condition
preparations	A list (concatenated and separated) of preparations and preservation methods for a specimen.	Optional	String	tissue (EDTA)
preparedDate	The date-time (in ISO 8601 format) representing the date the specimen was prepared..	Optional	Date (ISO)	2019-09-24
institutionCode	The name (or acronym) in use by the institution having custody of the object(s) or information referred to in the record.	Optional	String	StreamEnv
materialSampleID	An identifier for the MaterialSample (as opposed to a particular digital record of the material sample). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the materialSampleID globally unique	Optional	String	S07-06-01
associatedSequences	A list (concatenated and separated) of identifiers (publication, global unique identifier, URI) of genetic	Optional	String	https://www.ncbi.nlm.nih.gov/nuccore/MT766606.1

	sequence information associated with the Occurrence.			
verbatimIdentification	A string representing the taxonomic identification as it appeared in the original record.	Mandatory	String	Caladenia ?excelsa
dateIdentified	The date (in ISO 8601 format) on which the subject was determined as representing the Taxon.	Mandatory	Date (ISO)	2019-09-24
identifiedBy	A list (concatenated and separated) of names of people, groups, or organisations who assigned the Taxon to the subject.	Optional	String	J. Doe
scientificName	The full scientific name, with authorship and date information if known. When forming part of an Identification, this should be the name in lowest level taxonomic rank that can be determined. This term should not contain identification qualifications, which should instead be supplied in the IdentificationQualifier term.	Optional	String	Caladenia excelsa
scientificNameAuthorship	The authorship information for the scientificName formatted according to the conventions of the applicable nomenclaturalCode.	Optional	String	Hopper & A.P.BR
identificationQualifier	A brief phrase or a standard term ("cf.", "aff.") to express the determiner's doubts about the Identification.	Optional	String	Species incerta
identificationRemarks	Comments or notes about the Identification.	Optional	String	no flowers present
Kingdom	The full scientific name of the kingdom in which the taxon is classified.	Mandatory	String	Plantae
taxonRank	The taxonomic rank of the	Optional	String	species

	most specific name in the scientificName as it appears in the original record.			

Step 4: Request a secure upload link

When you are ready to upload your completed template, you can request a secure upload link. This link will be generated and sent to your nominated email address, and will be valid for 24 hours.

Step 5: See acknowledgement of your request

You will be directed to a web page confirming your request for that secure link.

Step 6: Receive email containing secure upload link

Check your email Inbox for the automated response from the BDR system with your link. You may need to check your Junk Mail or Spam folder if you do not receive it within a few minutes.

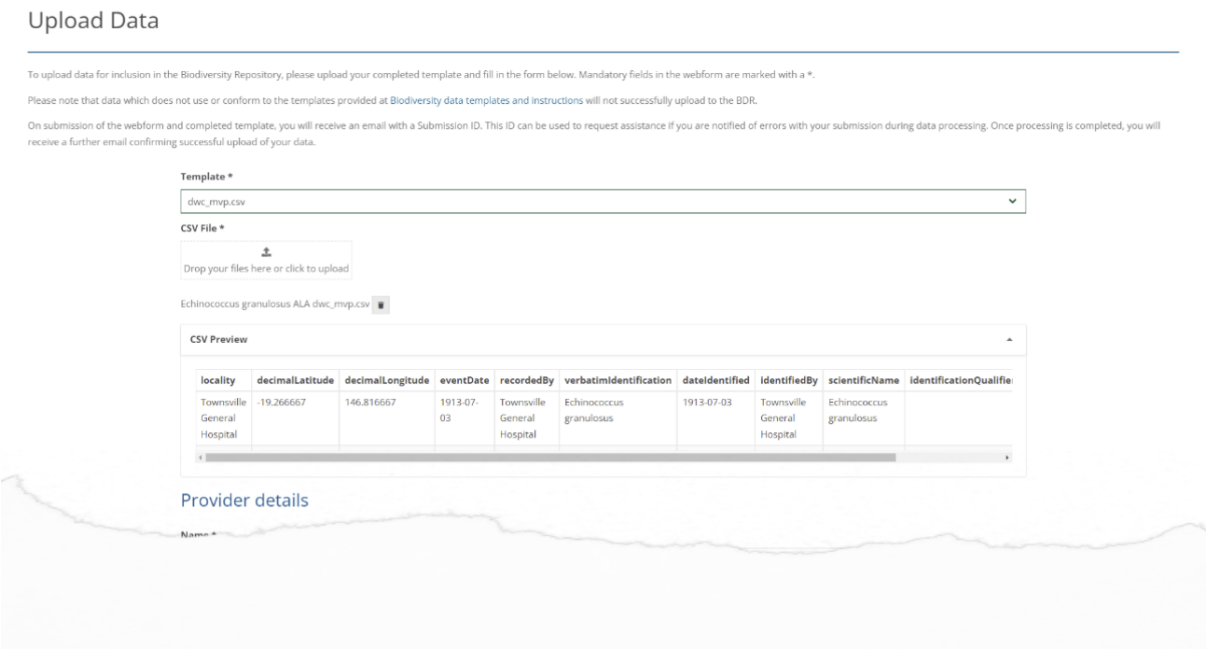
Step 7: Click on the secure upload link and open the webform

By clicking on the link provided in the email, you will be redirected to a webform, where you can upload your data file and fill in some details about your submission (described in Step 9).

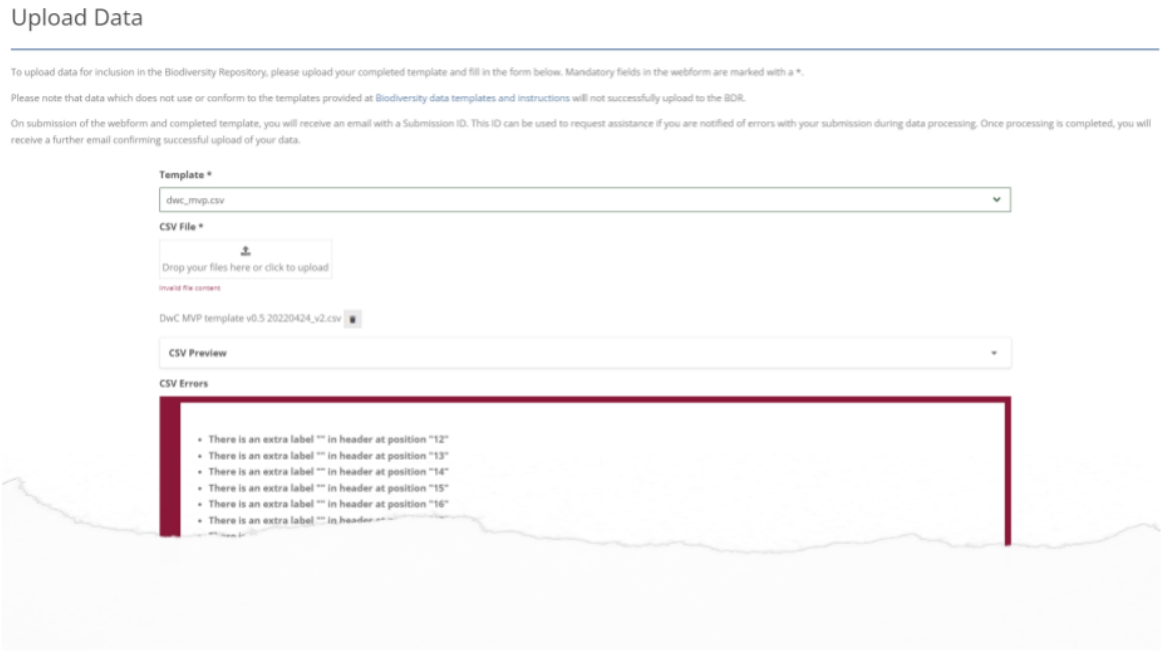
Step 8: Attach your data file in the completed template

Select the template from the list that matches your completed CSV. Then attach your completed CSV template to the webform, either through the ‘drag and drop’ method or by clicking on the upload button.

This will initiate a validation process that will check your data submission for format and completeness. If your file passes the validation process you will be presented with a preview that looks like the example below.



If there are validation issues with your dataset, a red preview of the errors will appear on screen to give you a sense of any errors that have been detected. Often, the same error will be repeated on multiple lines of your dataset.



The following are a list of reasons why your data might be rejected:

- Missing mandatory columns
- Missing record values for mandatory fields
- Incorrectly formatted field entries (e.g. incorrect date formats, text in a number field)
- Controlled vocabulary fields with values that do not exactly match a valid entry

- Latitude and Longitude that do not fit within an acceptable range. Refer to the Further Information section for more details.

The data format and completeness checks are not designed to check the content of your biodiversity data; another set of data quality and curation happens later in the process after your data is in the BDR system. The purpose of these initial checks is to immediately spot invalid or incomplete information, and prompt you to correct the errors for resubmission. It is critical, for instance, to ensure that all date entries in the BDR system comply with the ISO 8601 format (e.g. 2022-05-04), and that with fields subject to controlled vocabularies that no ambiguities exist from alternative or incorrect spelling.

Step 9: Fill in the rest of the webform

Once you have a CSV where the preview is showing as being free of validation errors, then fill out the fields in the webform. You could have completed step 9 first, but we recommend ensuring the data can actually be submitted before filling in the rest of the form.

In the webform you will be asked to fill out information about your submission, including the provider details, dataset details and any restrictions on data usage within the BDR. Mandatory fields in the webform are marked with an asterisk (*).

You will also be asked to acknowledge a licensing and privacy agreement, and to confirm the data will be made available to the public through a specified Creative Commons licence agreement.

Step 10: Click Submit

You are ready, and can click the Submit button! This will start processing your uploaded data.

Step 11: Receive email that upload has started

The processing of your data may take up to a few minutes, depending on the volume of data you are submitting. While this is in progress, you will receive an email notification that your submission is being processed.

Step 12: Receive email confirming successful or failed upload

On successful submission, the nominated email address will then receive a confirmation email with your Submission ID and a link to your data in the BDR system.

If at this point, if your submission still fails one of the checks above, you will receive an email at the nominated address with your Submission ID, the error code(s) encountered, and a short description on what to do next.

Further information

The valid latitude and longitude coordinates range in the BDR system is based on information from Geoscience Australia. The **Australian Continent** coordinate extremities can be found on this site:
https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/continental-extremities

The coordinates range for the seven **Australian External and Offshore territories** can be found on these sites:
https://d28rz98at9flks.cloudfront.net/70562/AustraliaAndExternalTerritories.pdf

https://www.ga.gov.au/scientific-topics/national-location-information/dimensions/remote-offshore-territories)

In summary:

- **Heard and McDonalds Island**
  Lat range: -52.902770 and -53.195018
  Long range: 73.872715 and 72.577376
- **Lord Howe Island**
  Lat range: -31.486129 and -31.787767
  Long range: 159.280368 and 159.036807
- **Macquarie Island**
  Lat range: -54.355874 and -55.123198
  Long range: 158.998625 and 158.674929
- **Norfolk Island**
  Lat range: -29.136568 and -28.994170
  Long range: 167.998035 and 167.913770
- **Ashmore and Cartier Islands**
  Lat range: -12.184700 and -12.547300
  Long range: 123.581854 and 122.927010
- **Christmas Island**
  Lat range: -10.412390 and -10.570559
  Long range: 105.712810 and 105.533149
- **Cocos Islands**
  Lat range: -11.822133 and -12.211000
  Long range: 96.930763 and 96.815497
- **Coral Sea Islands**
  Lat range: -15.721024 and -29.982747
  Long range: 159.140729 and 147.839456
- **Australian Antarctic Territory**
  Lat range: -64.928558 to -90.000026
  Long range: 167.724334 to 45.000000

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