

bdDwC User Guide

Authors: TBA

2018-09-03

Contents

Introduction	5
1 Installing bdDwC	7
1.1 Stable version from CRAN	7
1.2 Development version from GitHub	7
1.3 Possible problems & solutions	7
2 The shiny app	9
2.1 Launching the app	9
2.2 App overview	9
2.3 Data upload	9
2.4 Dictionaries	11
2.5 Darwinizing your dataset	11
2.6 Darwinizer results	11
2.7 Download your Darwinized data	14
2.8 Closing the app	14
2.9 References	14
3 Command line operations	15
3.1 Load package	15
3.2 Darwinizing a dataset	15
3.3 Updating the Darwin Cloud dictionary	15
4 Examples	17
5 References	19
6 Getting your feedback	21
6.1 Report a bug	21
6.2 Contribute	21
7 bdDwC citation	23
8 Learn more about Darwin Core	25
8.1 References	25

Introduction

[TBA]

Motivation

[TBA]

Chapter 1

Installing bdDwC

1.1 Stable version from CRAN

[Need To Be Updated!]

```
# The easiest way to get bdDwC is to install the whole bdverse:  
install.packages("bdverse")
```

```
# Alternatively, install just bdDwC:  
install.packages("bdDwC")
```

1.2 Development version from GitHub

Windows users install Rtools first.

```
install.packages("devtools")  
devtools::install_github("bd-R/bdDwC")
```

1.3 Possible problems & solutions

[TBA]

1.3.1 ???

TBA

1.3.2 ????

TBA

Chapter 2

The shiny app

2.1 Launching the app

```
library(bdDwC) # Upload package library
runDwC() # Launch the app
```

2.2 App overview

[Need To Be Updated!]

In the first screen, you'll need to upload or download your biodiversity data; choose dictionary and run the Darwinizer.

2.3 Data upload

2.3.1 From a local file

A CSV file or a Darwin Core Archive (DwC-A) zip file can be uploaded.

Local file size cannot exceed 1GB [?]

[Need To Be Updated!]

2.3.2 From an online database

Also, data can be retrieved directly from various online biodiversity databases. You need only to:

- Select the database
- Specify the desired scientific name.
- Specify the number of records (upper limit of 50,000).
- Check the box if records must have coordinates.
- Wait for data to be downloaded.

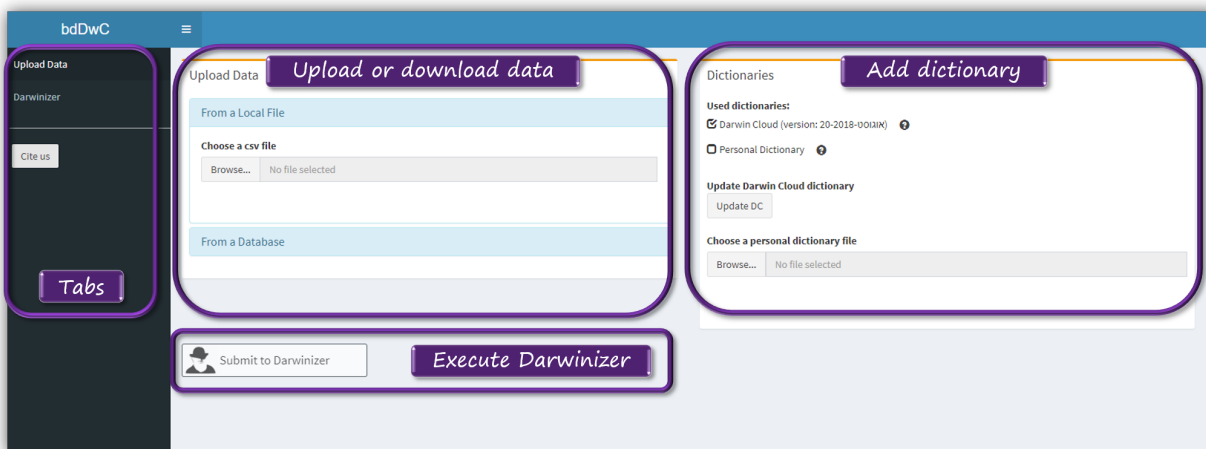


Figure 2.1: bdDwC App Overview

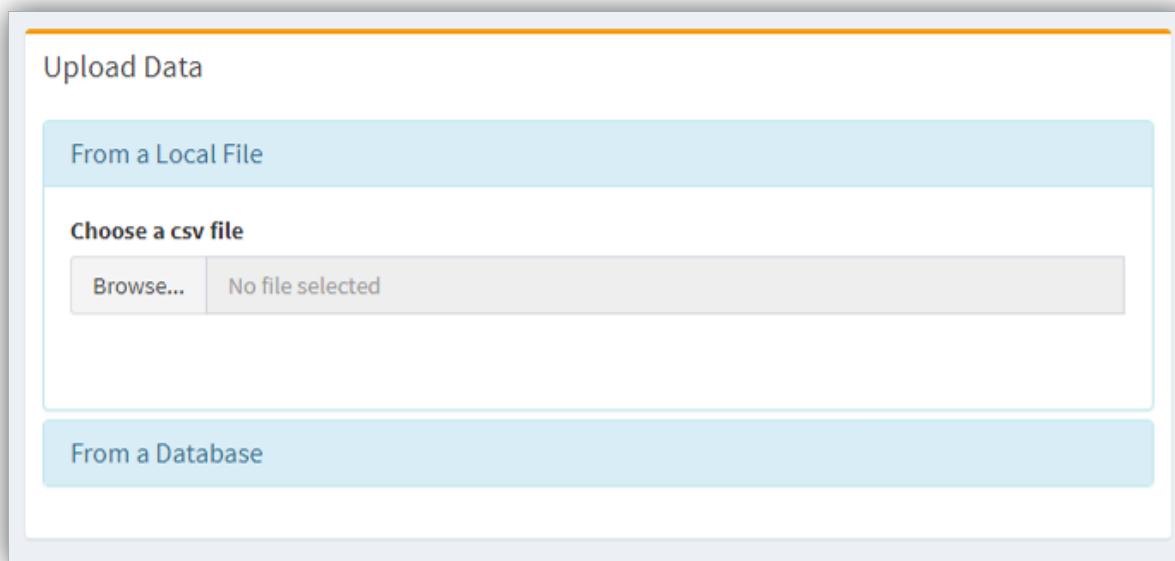


Figure 2.2: Data upload from a local file

[Need To Be Updated!]

2.4 Dictionaries

A dictionary is a key component when Darwinizing a dataset. It's basically a lookup table that lists a possible variation of field name and its corresponding DwC name.

2.4.1 The Darwin Cloud dictionary

The Darwin Cloud dictionary (Wieczorek et al., 2017), is a lookup table that accumulates different variations in DwC field names from different publishers. This valuable dictionary was created and is maintained by the Kurator project (<http://kurator.acis.ufl.edu/kurator-web/>), which provides workflow tools for data quality improvement of biodiversity data, via a user-friendly web interface.

Updating the Darwin Cloud

It's recommended to update the Darwin Cloud file. This can be done easily by clicking the **Update DC** button.

[Need To Be Updated!]

2.4.2 Your own dictionary

It's also possible to add your own dictionary by simply creating a CSV file with two columns, one for the Field Names and one for the Standard Names.

[Need To Be Updated!]

2.5 Darwinizing your dataset

Once a dataset is uploaded, the 'Submit to Darwinizer' button is activated. Clicking it will Darwinize the dataset.

[Need To Be Updated!]

2.6 Darwinizer results

2.6.1 Results page overview

[Need To Be Updated!]

Manually renaming field names can be done very easily, just choose the two corresponding fields and click the Rename button.

[Need To Be Updated!]

Hovering over a DwC standard name will display its description.

Upload Data

From a Local File

From a Database

Scientific Name:

Puma concolor

Record Size:

500 50,000

0 5,000 10,000 15,000 20,000 25,000 30,000 35,000 40,000 45,000 50,000

Online Database:

☒ GBIF

☐ Bison

☐ Inat

☐ eBird

☐ Ecoengine

☐ Vertnet

Query Database

Figure 2.3: Data upload from online biodiversity databases

Update Darwin Cloud dictionary

Update DC

Figure 2.4: Updating the Darwin Cloud

Choose a personal dictionary file

Browse... dictionary_for-ALA.csv

Upload complete

Select field and standard names

Field Names

☒ fieldname

☐ |..standard.name

Standard Names

☐ fieldname

☒ |..standard.name

Figure 2.5: Uploading your own dictionary

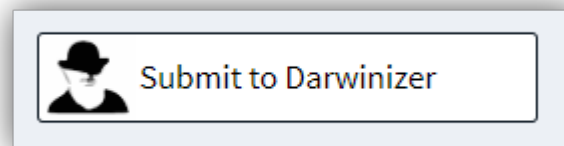


Figure 2.6: Submit to Darwinizer button

bdDwC

Upload Data

Darwinizer

Cite us

Statistics

47 Names Submitted	5 (11%) Names Darwinized	4 Darwinized: Matched	0 Darwinized: Manually	1 Darwinized: Identical
-----------------------	-----------------------------	--------------------------	---------------------------	----------------------------

Action buttons

Manual renaming

User Names

- ☒ Basis.Of.Record...original
- ☐ Catalog.Number
- ☐ Class...matched
- ☐ Collection.Code
- ☐ Coordinate.Uncertainty.in.Metres...parsed
- ☐ Country...parsed
- ☐ Data.Resource.ID
- ☐ Data.Resource.Name
- ☐ Family...matched
- ☐ Genus...matched
- ☐ geodetic.Datum
- ☐ IBRA.7.Regions
- ☐ IMCRA.4.Regions
- ☐ Institution.Code
- ☐ Kingdom...match
- ☐ Latitudes...original

Stand Names

- ☒ acceptedNameUsage
- ☐ acceptedNameUsageID
- ☐ accessRights
- ☐ associatedMedia
- ☐ associatedOccurrences
- ☐ associatedOrganisms
- ☐ associatedReferences
- ☐ associatedSequences
- ☐ associatedTaxa
- ☐ basisOfRecord
- ☐ bed
- ☐ behavior
- ☐ biographicCitation
- ☐ catalogNumber

Name conversion results

Darwinized Names

- ☐ Basis.Of.Record...processed -> 1
- ☐ Collector -> 19
- ☐ Event.Date...parsed -> 2
- ☐ Sex -> 151

Manually Renamed

Nothing was renamed

Identical matches

- ☐ locality -> locality

Darwinized

Manual

Identical

Figure 2.7: Darwinizer results

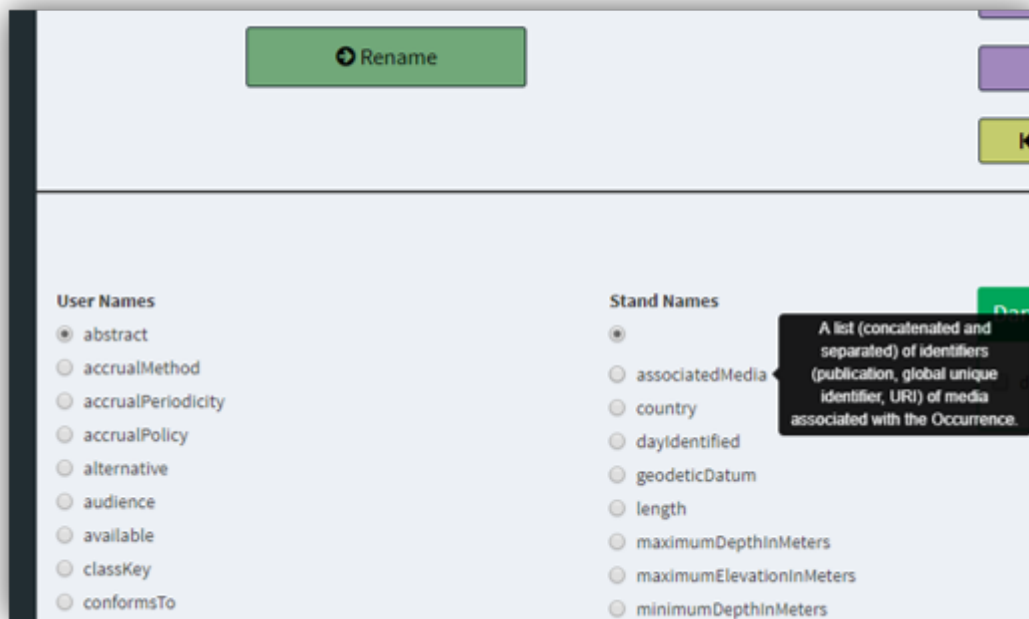


Figure 2.8: Manually renaming fields

2.7 Download your Darwinized data

[Need To Be Updated!]

2.8 Closing the app

2.9 References

Chapter 3

Command line operations

3.1 Load package

Load the bdDwC package

```
library(bdDwC)
```

3.2 Darwinizing a dataset

bdDwC contains Indian Reptile dataset `bdDwC:::dataReptiles`.

The function to Darwinize a dataset is `darwinizeNames` (replace `bdDwC:::dataReptiles` with wanted dataset):

```
result <- darwinizeNames(dataUser = bdDwC:::dataReptiles,  
                        dataDWC   = bdDwC:::dataDarwinCloud$data)
```

You can replace `bdDwC:::dataReptiles` with your dataset

Rename your dataset field names to Darwinized names using `renameUserData`:

```
renameUserData(bdDwC:::dataReptiles, result)
```

3.3 Updating the Darwin Cloud dictionary

To get newest version of Darwin Cloud Data run:

```
downloadCloudData()
```

which will download data from the remote repository and extract field and standard names.

Chapter 4

Examples

[**TBA**]

I'm thinking to show the Darwinizer app with four distinct datasets:

1. ALA dataset in DwC format
2. ALA dataset in legacy format
3. VertNet dataset
4. GBIF dataset

Chapter 5

References

Chapter 6

Getting your feedback

Loading...

6.1 Report a bug

Submit an issue at <https://github.com/bd-R/bdDwC/issues>

6.2 Contribute

Contribute: <https://github.com/bd-R/bdDwC>

Join: <https://bd-r-group.slack.com>

Chapter 7

bdDwC citation

```
citation("bdDwC")
```

```
##
## To cite package 'bdDwC' in publications use:
##
## Povilas Gibas, Tomer Gueta, Vijay Barve, Thiloshon Nagarajah and
## Yohay Carmel (2018). bdDwC: Darwinizer: Darwin Core (DwC) Field
## Names Standardization. R package version 0.1.15.
##
## A BibTeX entry for LaTeX users is
##
## @Manual{,
##   title = {bdDwC: Darwinizer: Darwin Core (DwC) Field Names Standardization},
##   author = {Povilas Gibas and Tomer Gueta and Vijay Barve and Thiloshon Nagarajah and Yohay Carmel},
##   year = {2018},
##   note = {R package version 0.1.15},
## }
```


Chapter 8

Learn more about Darwin Core

-
- Darwin Core Hour webinar series
 - The Darwin Core Questions & Answers wiki
 - GBIF: What is Darwin Core, and why does it matter?
 - Darwin Core: An Evolving Community-Developed Biodiversity Data Standard (Wieczorek et al., 2012)

8.1 References

Bibliography

- Wieczorek, J., Bloom, D., Guralnick, R., Blum, S., Döring, M., Giovanni, R., Robertson, T., and Vieglaiss, D. (2012). Darwin core: An evolving community-developed biodiversity data standard. *PLOS ONE*, 7(1):1–8.
- Wieczorek, J., Morris, P. J., Hanken, J., B. Lowery, D., Ludäscher, B., Macklin, J., McPhillips, T., A. Morris, R., and Zhang, Q. (2017). Darwin cloud: Mapping real-world data to darwin core. *Biodiversity Information Science and Standards*, 1:e20486.