`bdDwC` User Guide

Authors: TBA

2018-09-03

Contents

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) # Uplaod 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.

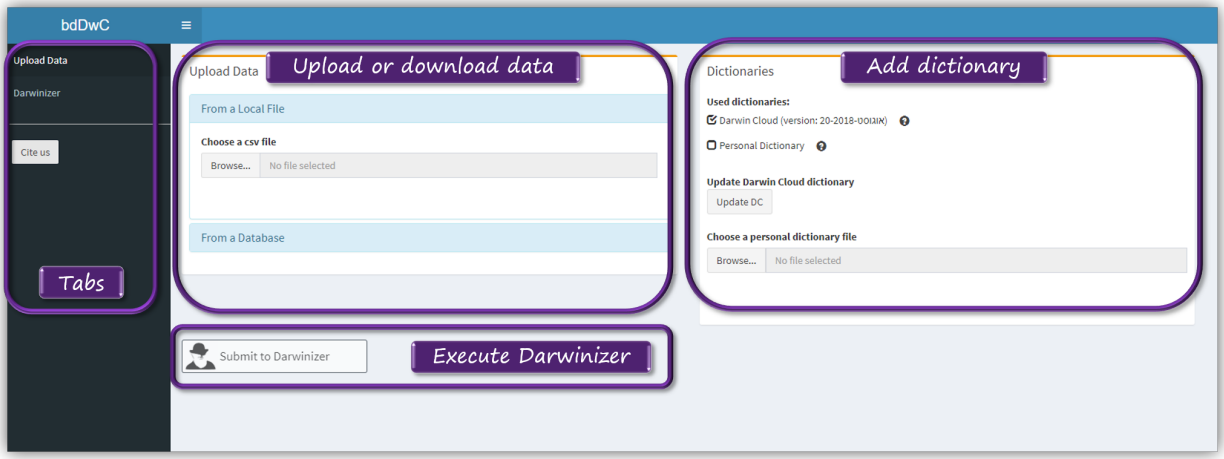

Figure 2.1: bdDwC App Overview

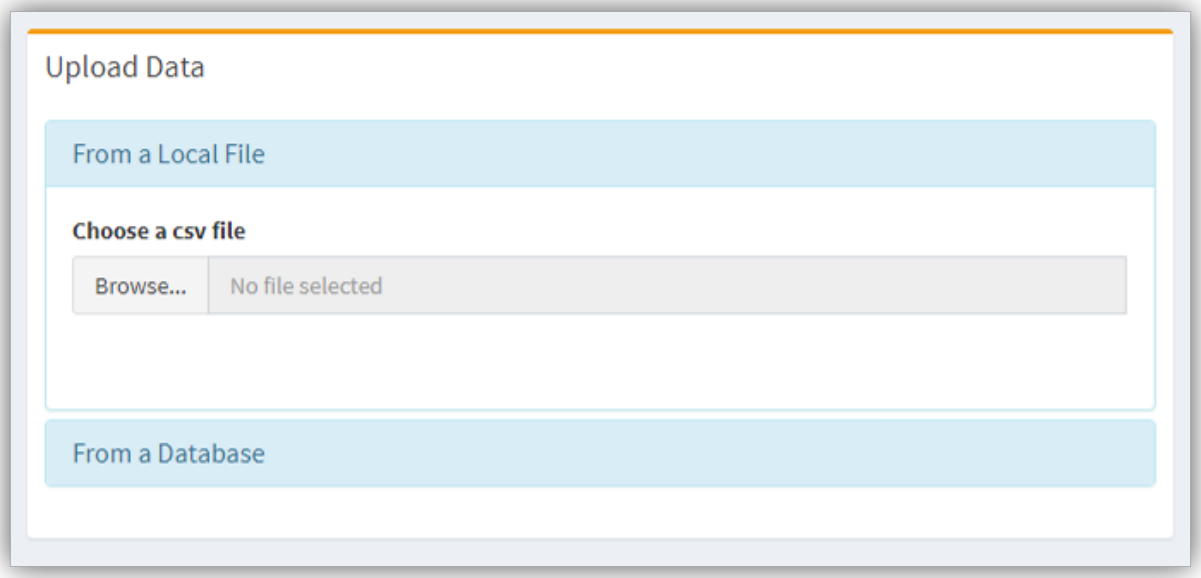

Figure 2.2: Data upload from a local file

[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 it 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 overwiew

[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.

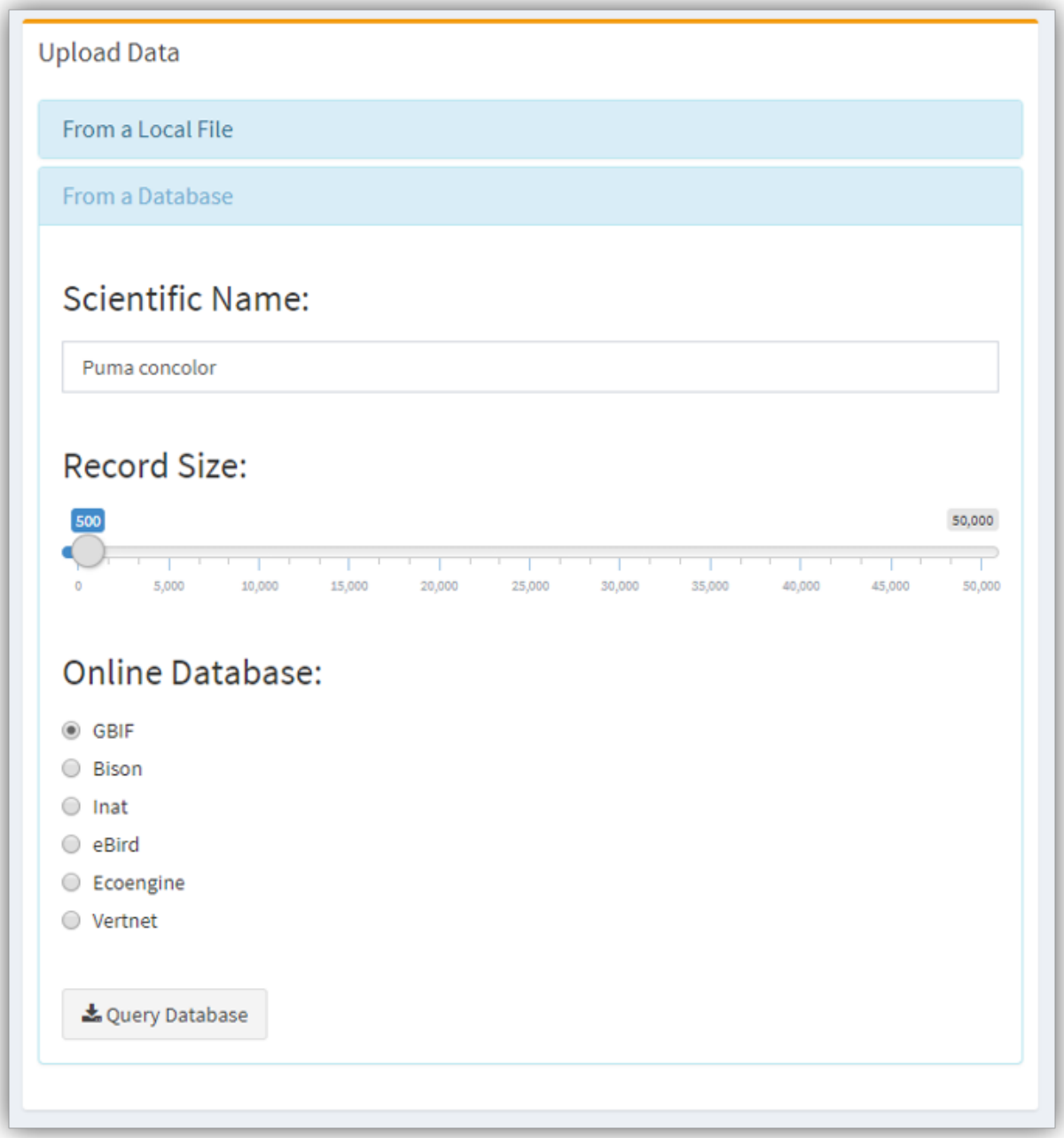

Figure 2.3: Data upload from online biodiversity databases

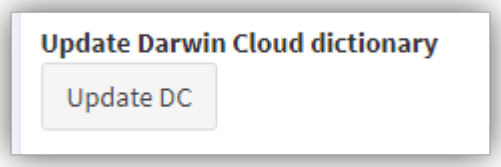

Figure 2.4: Updating the Darwin Cloud

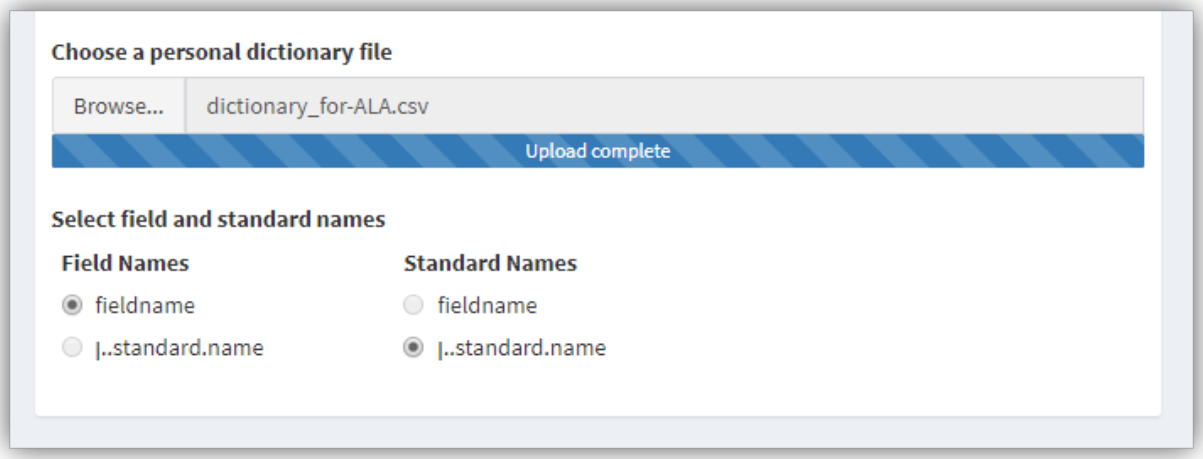

Figure 2.5: Uploading your own dictionary

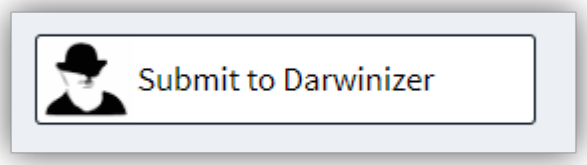

Figure 2.6: Submit to Darwinizer button

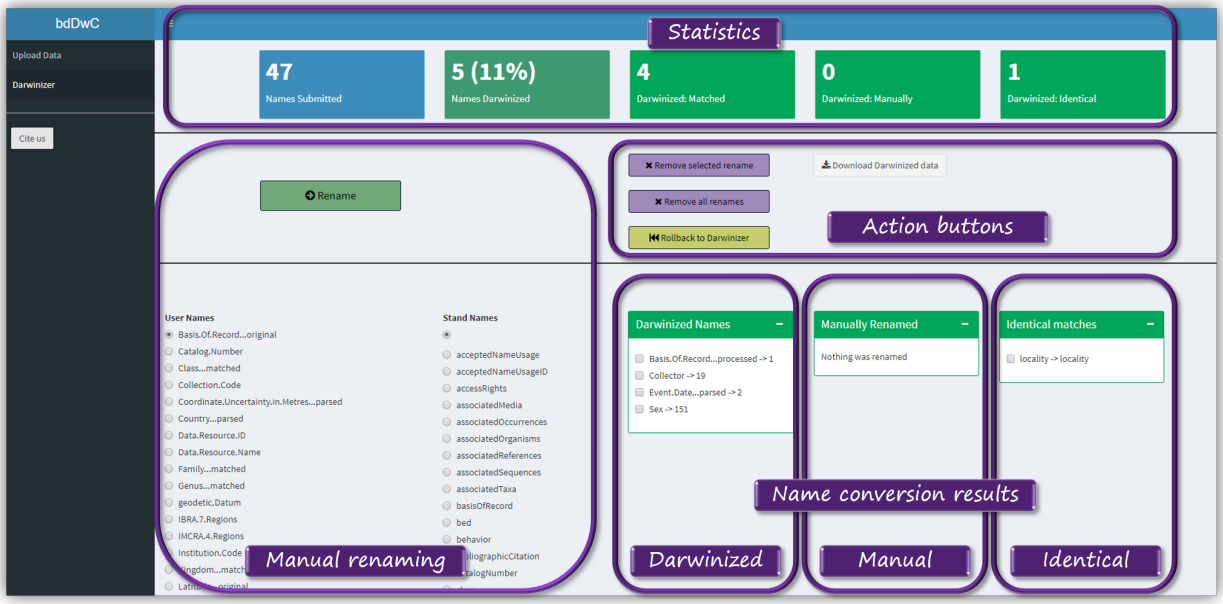

Figure 2.7: Darwinizer results

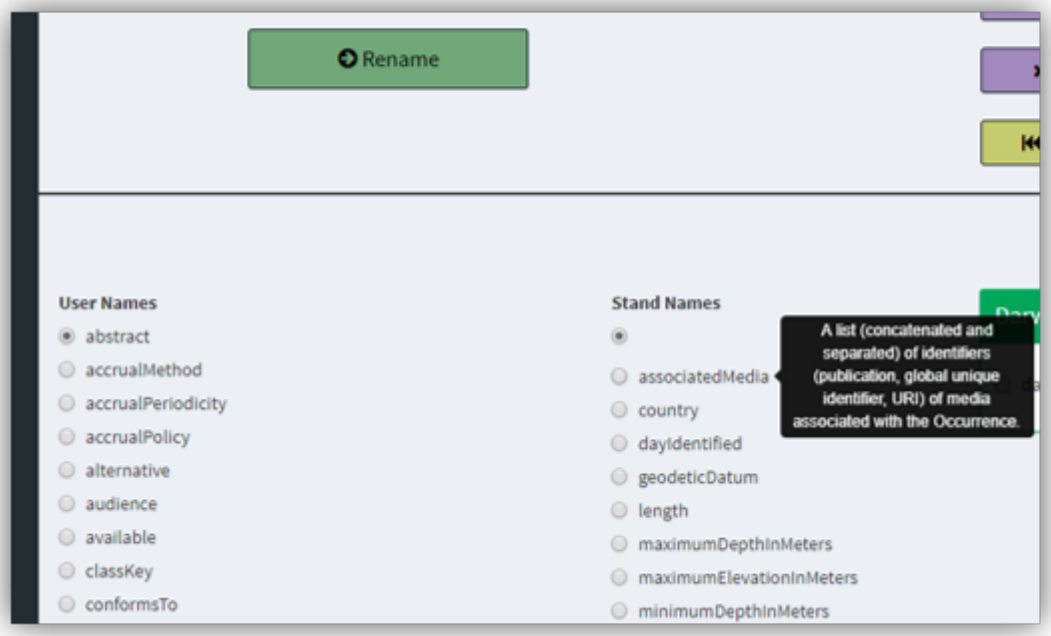

Figure 2.8: Manually renaming fields

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 isdarwinizeNames (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 Vieglais, 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.