`bdDwC` User Guide

Authors: Tomer Gueta and Povilas Gibas

built on 2018-10-15 - for bdDwC v0.1.20

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 . . . 12
2.5 Darwinizing your dataset . . . 13
2.6 Darwinizer results . . . 13
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 Getting your feedback **19**
5.1 Report a bug . . . 19
5.2 Contribute . . . 19

6 `bdDwC` citation **21**

7 Learn more about Darwin Core **23**

Introduction

`bdDwC` is an R package that supplies a Shiny app and a set of functions for standardizing field names according to the Darwin Core (DwC) format. `bdDwC` is a key element in the `bdverse`– a collection of tools, that form a general framework for facilitating biodiversity science in R.

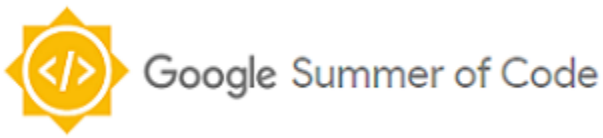
What is the Darwin Core standard?

Darwin Core (DwC) is a global standard for publishing biodiversity data, whose goal is to facilitate the sharing of biodiversity information, by providing identifiers, labels, and definitions (Wieczorek et al., 2012). DwC was established as an evolving community-developed standard, by the Biodiversity Information Standards Working Group (www.tdwg.org). DwC is a library of definitions of common biodiversity data terms, each of which represents a field within the database. There are around 200 such fields (not including DwC extensions); a full set of the DwC terms with their descriptions is available in the Quick Reference Guide (http://rs.tdwg.org/dwc/terms). For more information see section 6.

Why it's important to "Darwinize" a dataset

Running the Darwinizer enables you to standardize many field names in your dataset – and that allows the `bdverse` to handle data from various biodiversity portals, and lets you enjoy all of `bdvers` features, regardless of publishers variation in field names.

Fundings



See the GSoC project idea page

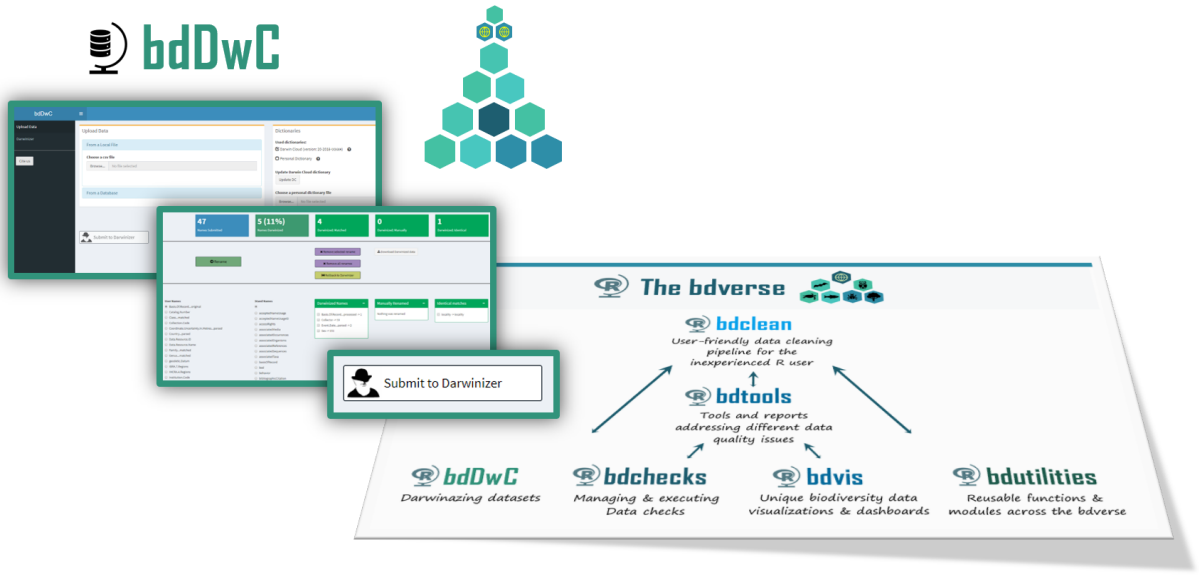


Figure 1: bdDwC in the bdverse

This work is supported by the Israel Science Foundation (ISF) Grant No. 127/16

Figure 2:

Chapter 1

Installing bdDwC

1.1 Stable version from CRAN

```
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

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.

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.

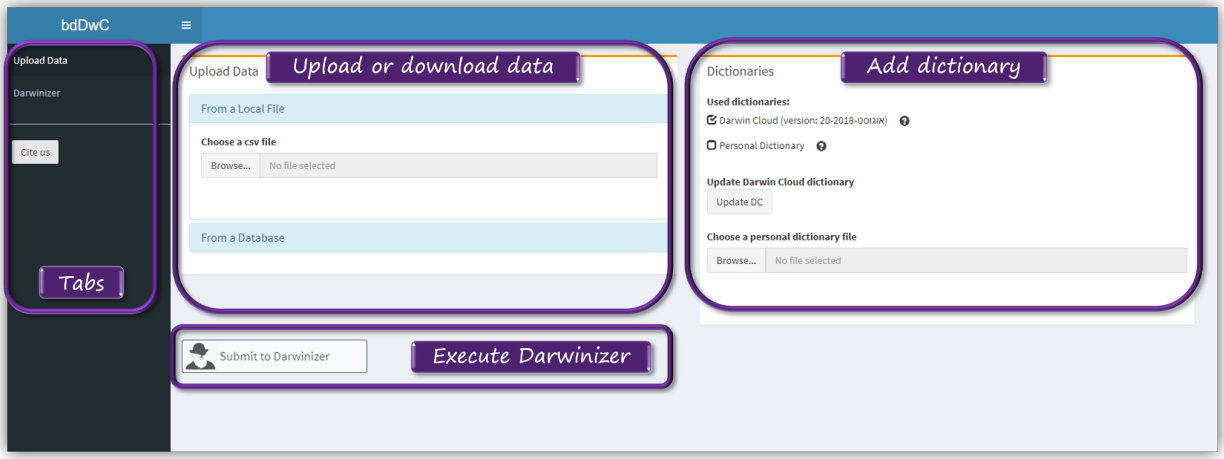


Figure 2.1: bdDwC App Overview

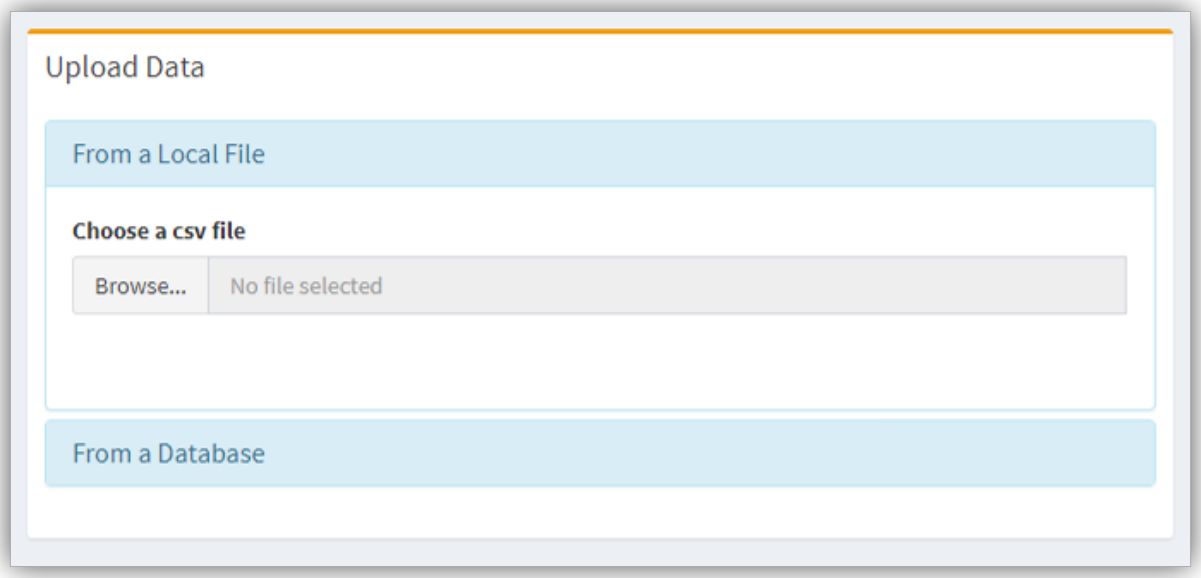


Figure 2.2: Data upload from a local file

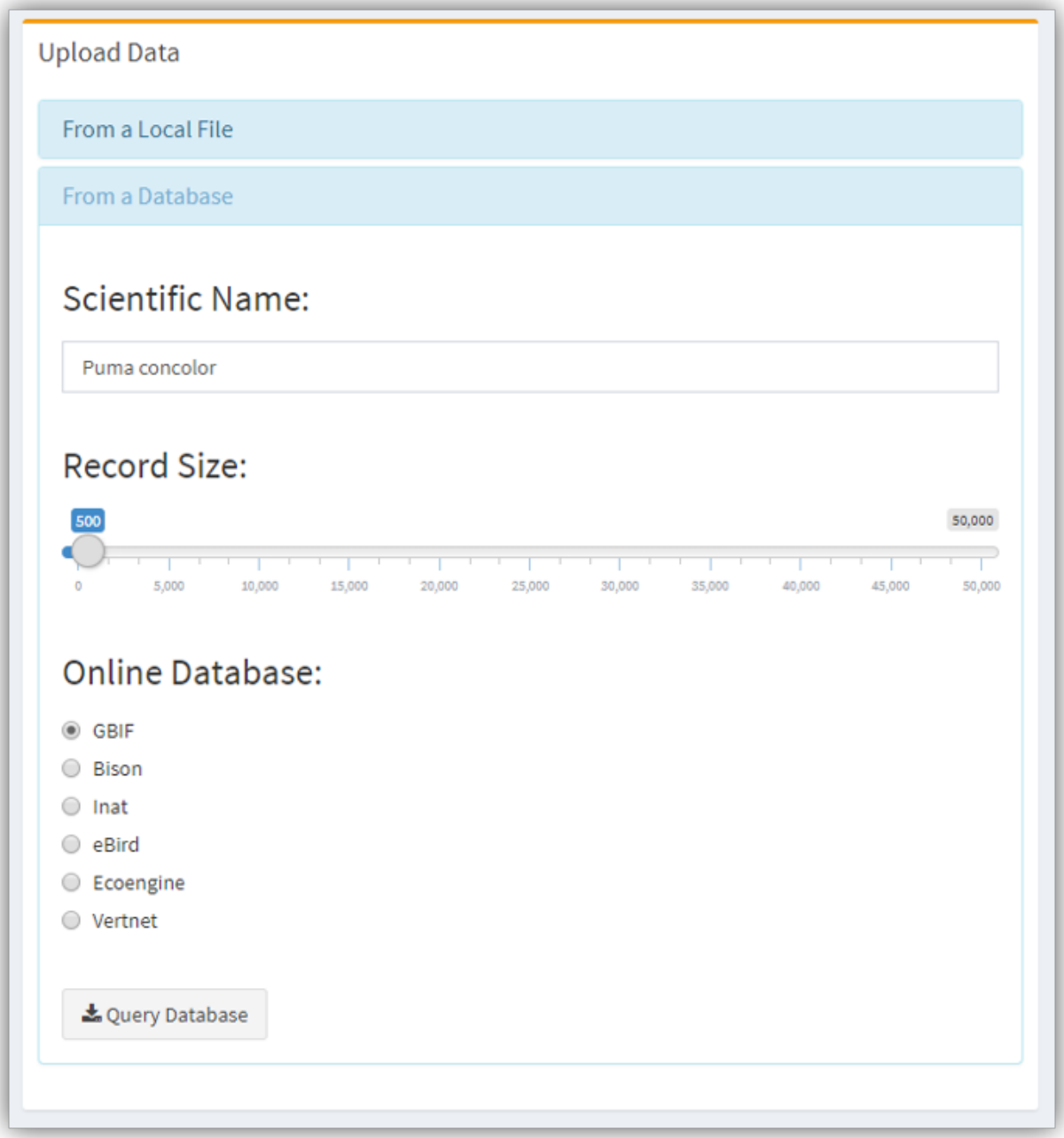


Figure 2.3: Data upload from online biodiversity databases

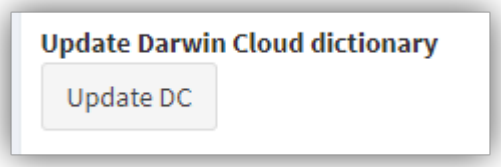


Figure 2.4: Updating the Darwin Cloud

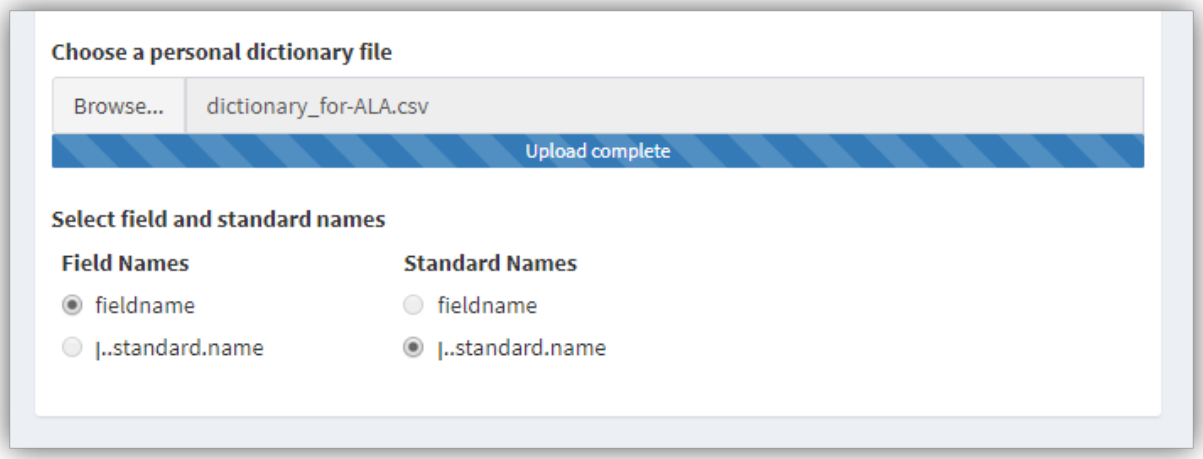


Figure 2.5: Uploading your own dictionary

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 and critical 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. The development of bdDwC was inspired by Kurator's own Darwinizer.

Updating the Darwin Cloud

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

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.

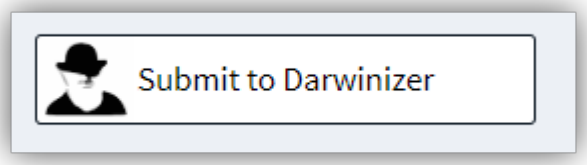


Figure 2.6: Submit to Darwinizer button

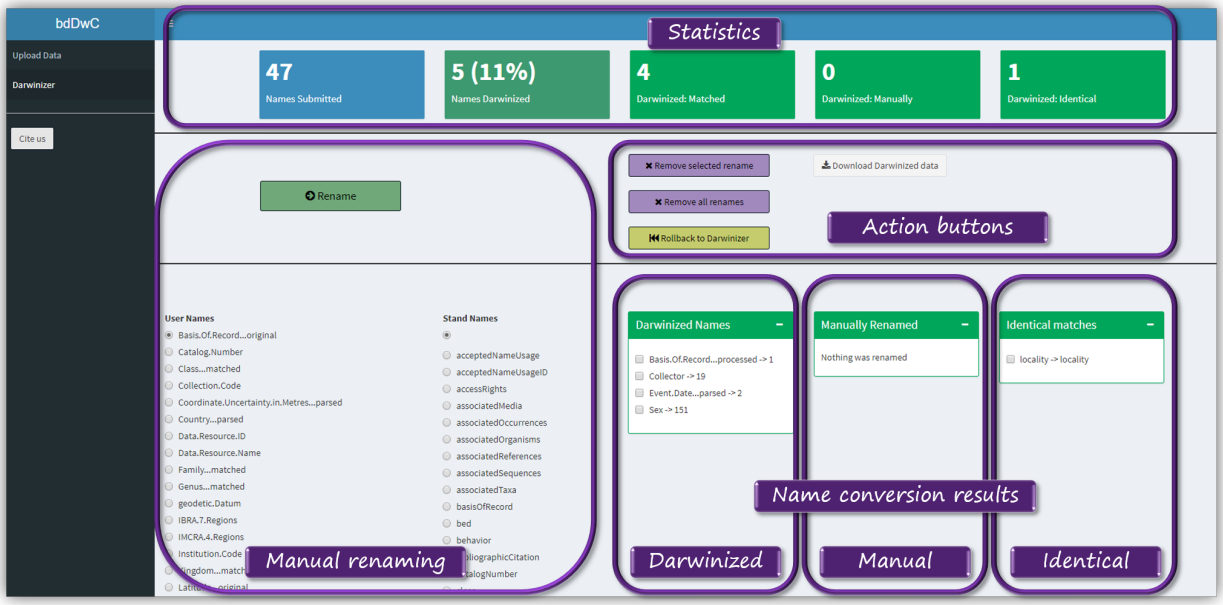


Figure 2.7: Darwinizer results

2.5 Darwinizing your dataset

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

2.6 Darwinizer results

2.6.1 Results page overwiew

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

Hovering over a DwC standard name will display its description.

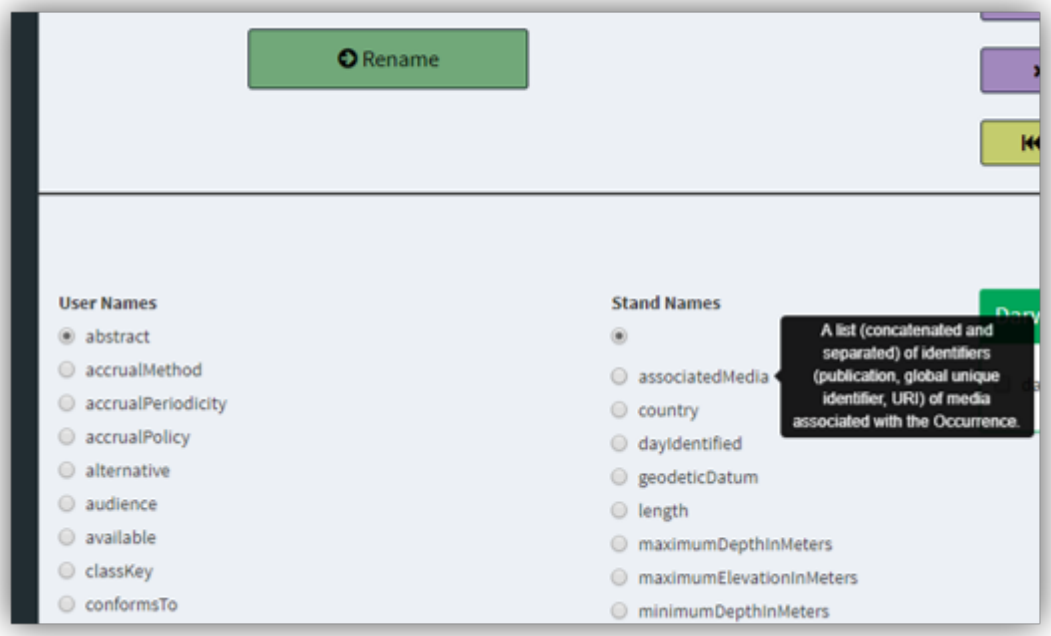


Figure 2.8: Manually renaming fields

2.7 Download your Darwinized data

2.8 Closing the app

Just close the app browser tab, and the R session will be terminated. To reopen it run in the R Console `runDwC()`.

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**]

Chapter 5

Getting your feedback

Loading...

5.1 Report a bug

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

5.2 Contribute

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

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

Chapter 6

bdDwC citation

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

```
##
## To cite package 'bdDwC' in publications use:
##
##   Povilas Gibas, Tomer Gueta, Vijay Barve, Thiloshon Nagarajah and
##   Yohay Carmel (2018). bdDwC: field names conversion to Darwin
##   Core (DwC) format. R package version 0.1.20.
##   https://github.com/bd-R/bdDwC
##
## A BibTeX entry for LaTeX users is
##
##   @Manual{,
##     title = {bdDwC: field names conversion to Darwin Core (DwC) format},
##     author = {Povilas Gibas and Tomer Gueta and Vijay Barve and Thiloshon Nagarajah and Yohay Carmel
##     year = {2018},
##     note = {R package version 0.1.20},
##     url = {https://github.com/bd-R/bdDwC},
##   }
```

Chapter 7

Learn more about Darwin Core

- **The Darwin Core Questions & Answers Site**
- **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)**

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):e29715.

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.