

The `dwctaxon` R package for editing and validating taxonomic data



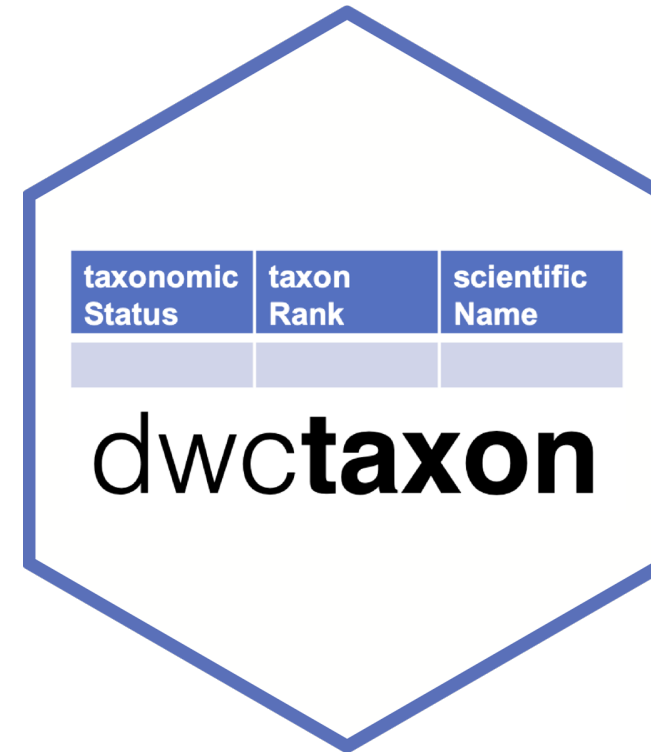
Joel H. Nitta¹

Wataru Iwasaki²

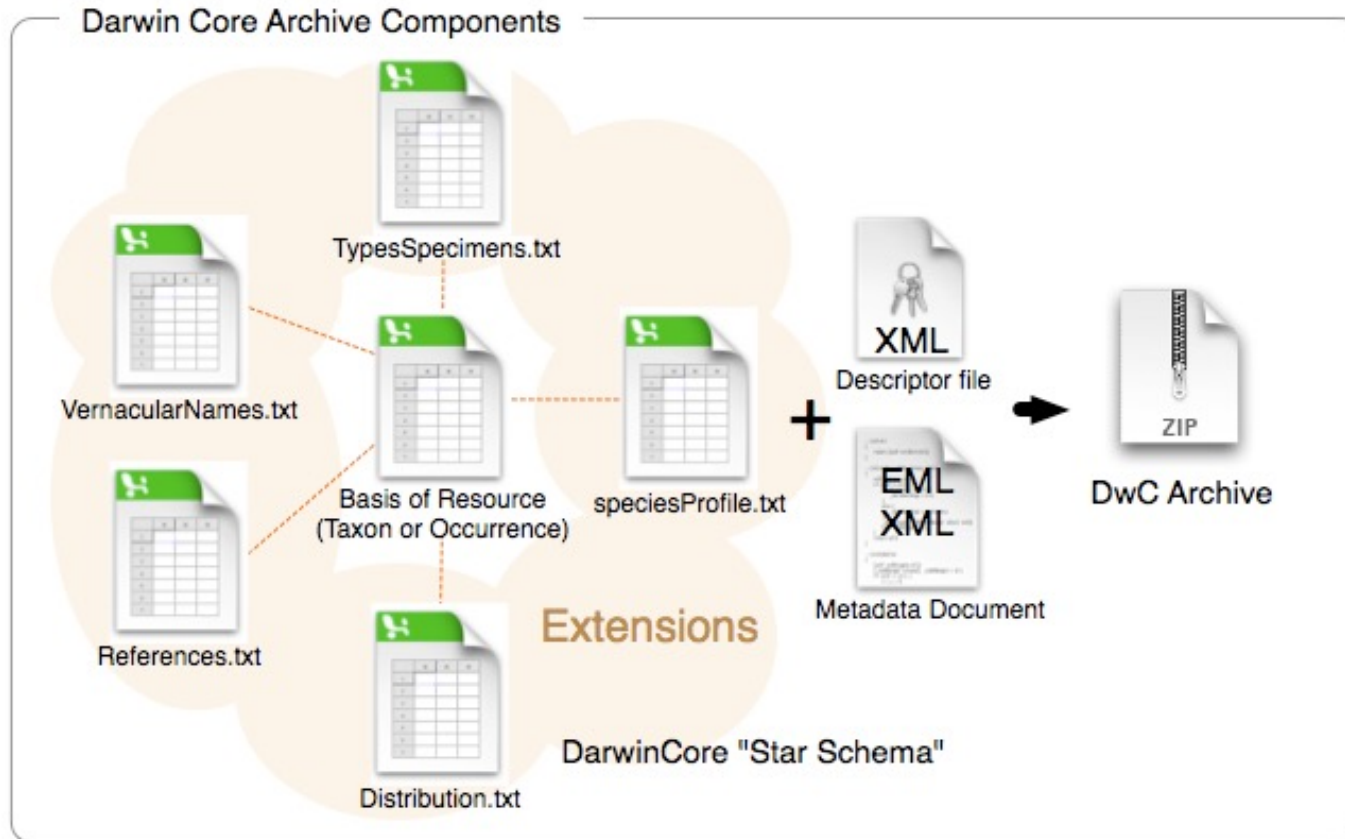
¹Chiba University, Japan

²The University of Tokyo, Japan

<https://www.joelnitta.com>



Darwin Core (DwC) is a standard for biodiversity data



dwc taxon helps edit and validate DwC taxonomic data

- Use cases
 - Taxonomic database management
 - Taxon checklists
 - Name standardization



The screenshot shows the top portion of the Canadensys website. At the top center is a circular logo with a globe and arrows. Below it, the word "Canadensys" is written in a large, white, sans-serif font. Underneath the logo and name is a dark navigation bar with several items: "explorer" with a dropdown arrow, "repository" in red text, "tools", and "vascan". Below the navigation bar, the text "Database of Vascular Plants of Canada (VASCAN)" is displayed in bold. A smaller line of text below that reads "Latest version published by Université de Montréal Biodiversity Centre on Apr 18, 2023". To the left of the main text is a green logo consisting of four interlocking squares. To the right of this logo, a paragraph of text describes VASCAN as a comprehensive and curated checklist of all vascular plants reported in Canada, Greenland (Denmark), and Saint Pierre and Miquelon (France). It also mentions that VASCAN was developed at the Université de Montréal Biodiversity Centre and is maintained by a

Example: Vascular Plants of Canada

<http://data.canadensys.net/ipt/resource.do?r=vascan>

Editing taxonomic data is complicated!

Old version

- Accepted species 1: *Cephalomanes crassum*
 - Synonym: *Trichomanes crassum*
- Accepted species 2: *Cephalomanes densinervium*
 - Synonym: *Trichomanes densinervium*

New version (*C. crassum* → synonym of ***C. densinervium***)

- Accepted species: ***Cephalomanes densinervium***
 - Synonym 1: *Cephalomanes crassum*
 - Synonym 2: *Trichomanes crassum*
 - Synonym 3: *Trichomanes densinervium*

Need to account for **all** synonyms!

dct_change_status() handles synonym mapping automatically

```
1 # C. crassum → synonym of C. densinervium
2 dct_change_status(
3   tax_dat = filmies,
4   sci_name = "Cephalomanes crassum (Copel.) M. G. Price",
5   new_status = "synonym",
6   usage_name = "Cephalomanes densinervium (Copel.) Copel."
7 )
```

```
# A tibble: 4 × 4
```

	taxonID	acceptedNameUsageID	taxonomicStatus	scientificName
	<chr>	<chr>	<chr>	<chr>
1	54133783	54115098	synonym	Trichomanes crassum Copel.
2	54115097	54115098	synonym	Cephalomanes crassum (Copel.) M....
3	54133784	54115098	synonym	Trichomanes densinervium Copel.
4	54115098	<NA>	accepted name	Cephalomanes densinervium (Copel...

Taxonomic data can be full of errors!

- Non-unique names
- Mistaken ranks
- Incorrect synonym mapping
- Etc...



https://media.npr.org/assets/img/2018/10/17/gettyimages-528623386_slide-564362f49c9eec2b225c747098405186a740c536-s600-c85.webp

dct_validate() finds problems in taxonomic data

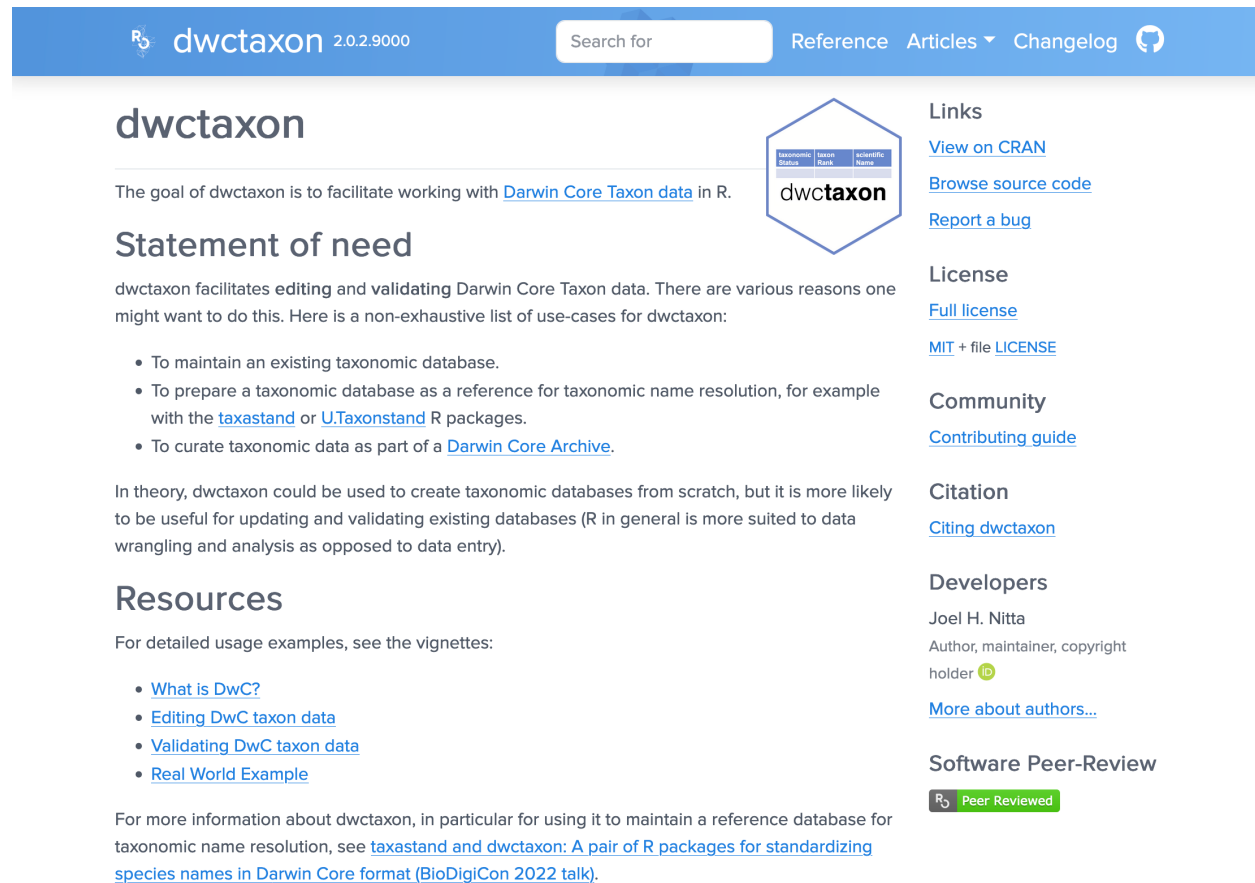
```
1  dct_change_status(  
2    tax_dat = filmies,  
3    sci_name = "Trichomanes crassum Copel.",  
4    new_status = "synonym",  
5    usage_name = "Trichomanes densinervium Copel."  
6  ) |>  
7  dct_validate()
```

```
Error: `check_mapping` failed.  
`taxonID`(s) detected whose `acceptedNameUsageID` value does not map to  
`taxonID` of an existing name.  
Bad `taxonID`: 54133783  
Bad `scientificName`: Trichomanes crassum Copel.
```

Please give `dwctaxon` a try!

- <https://github.com/ropensci/dwctaxon>

Thank you!



The screenshot shows the documentation page for the `dwctaxon` R package. The page has a blue header with the package name and version (2.0.2.9000), a search bar, and navigation links for Reference, Articles, Changelog, and GitHub. The main content area is divided into sections: a description of the package's goal, a Statement of need, Resources, and a Citation section. A sidebar on the right contains links for View on CRAN, Browse source code, Report a bug, License (Full license, MIT + file LICENSE), Community (Contributing guide), Citation (Citing dwctaxon), Developers (Joel H. Nitta), and Software Peer-Review (Peer Reviewed).

dwctaxon 2.0.2.9000

Search for Reference Articles Changelog

dwctaxon

The goal of `dwctaxon` is to facilitate working with [Darwin Core Taxon data](#) in R.

Statement of need

`dwctaxon` facilitates editing and validating Darwin Core Taxon data. There are various reasons one might want to do this. Here is a non-exhaustive list of use-cases for `dwctaxon`:

- To maintain an existing taxonomic database.
- To prepare a taxonomic database as a reference for taxonomic name resolution, for example with the [taxastand](#) or [U.Taxonstand](#) R packages.
- To curate taxonomic data as part of a [Darwin Core Archive](#).

In theory, `dwctaxon` could be used to create taxonomic databases from scratch, but it is more likely to be useful for updating and validating existing databases (R in general is more suited to data wrangling and analysis as opposed to data entry).

Resources

For detailed usage examples, see the vignettes:

- [What is DwC?](#)
- [Editing DwC taxon data](#)
- [Validating DwC taxon data](#)
- [Real World Example](#)

For more information about `dwctaxon`, in particular for using it to maintain a reference database for taxonomic name resolution, see [taxastand and dwctaxon: A pair of R packages for standardizing species names in Darwin Core format \(BioDigiCon 2022 talk\)](#).

Links

- [View on CRAN](#)
- [Browse source code](#)
- [Report a bug](#)

License

- [Full license](#)
- [MIT](#) + file [LICENSE](#)


Community

- [Contributing guide](#)

Citation


- [Citing dwctaxon](#)

Developers

Joel H. Nitta
Author, maintainer, copyright holder 

[More about authors...](#)

Software Peer-Review

 Peer Reviewed