

# 分類学データを正確・速やかに管理するRパッケージ 「dwctaxon」の紹介

ニッタ ジョエル<sup>1</sup>

岩崎 渉<sup>2</sup>

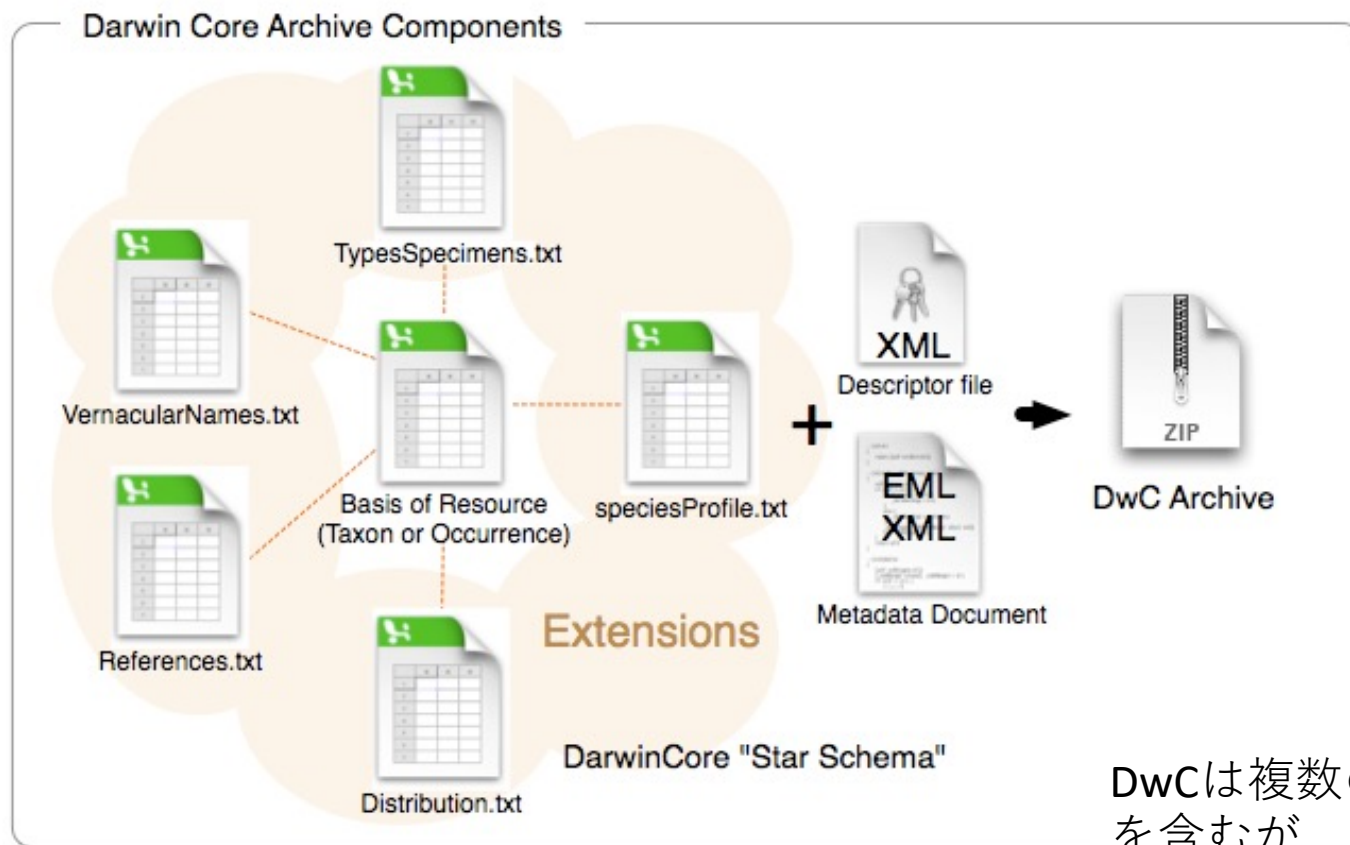
<sup>1</sup>千葉大学 大学院国際学術研究院

<sup>2</sup>東京大学 大学院新領域創成科学研究科

<https://www.joelnitta.com>



# Darwin Core (DwC) とは、生物多様性データの標準である



様々なデータベースに広く使われている (GBIFなど)

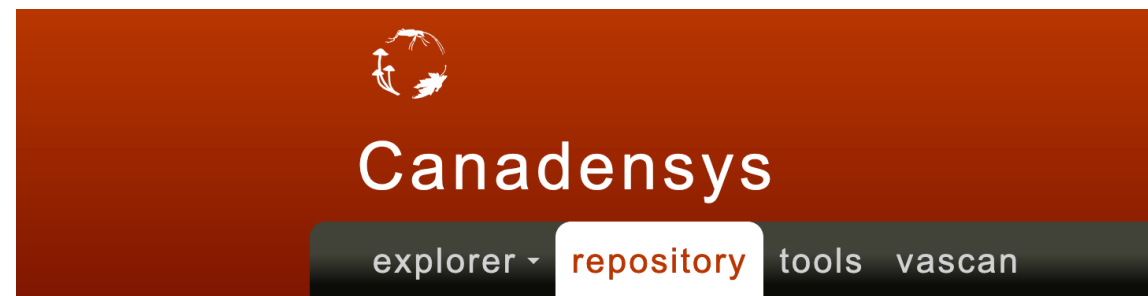


標準を使うことによって、データベース間のデータの共有ができる

DwCは複数のデータ型 (分布データなど) を含むが、今回の発表では**分類学的データ (種名)** だけを扱う

# dwctaxonとは、DwCデータの編集と検証を行うRパッケージである

- 利用の実例
  - 分類学的データベースの管理
  - チェックリストの管理
  - 種名の標準化



## Database of Vascular Plants of Canada (VASCAN)

*Latest version published by Université de Montréal Biodiversity Centre on Apr 18, 2023*



The Database of Vascular Plants of Canada or VASCAN (<http://data.canadensys.net/vascan>) is a comprehensive and curated checklist of all vascular plants reported in Canada, Greenland (Denmark), and Saint Pierre and Miquelon (France). VASCAN was developed at the Université de Montréal Biodiversity Centre and is maintained by a

カナダの維管束植物の [データベース](#)

dwctaxonで解析した例：

<https://docs.ropensci.org/dwctaxon/articles/real-data.html>

# 分類学的データの管理が想像以上に大変

例：前のバージョン

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

新しいバージョン (C. crassum → **C. densinervium**のシノニムに変える)

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

**全てのシノニムを考慮しないといけない！**

# dct\_change\_status() が自動的に シノニムなどを管理してくれる

```
1 # Original data: two accepted names, each with a synonym
```

```
2 filmies
```

```
# A tibble: 4 × 5
```

	taxonID	acceptedNameUsageID	taxonomicStatus	taxonRank	scientificName
	<chr>	<chr>	<chr>	<chr>	<chr>
1	54133783	54115097	synonym	species	Trichomanes crassum Co...
2	54115097	<NA>	accepted	species	Cephalomanes crassum (...
3	54133784	54115098	synonym	species	Trichomanes densinervi...
4	54115098	<NA>	accepted	species	Cephalomanes densinervi...

# dct\_change\_status() が自動的に シノニムなどを管理してくれる

```
1 # Change C. crassum to synonym of C. densinervium
2 filmies |>
3   dct_modify_row(
4     scientificName = "Cephalomanes densinervium (Copel.) Copel.",
5     taxonomicStatus = "synonym",
6     acceptedNameUsage = "Cephalomanes crassum (Copel.) M. G. Price",
7     stamp_modified = FALSE
8   )
```

```
# A tibble: 4 × 4
```

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

# 分類学的データはエラーだらけのことがある！

- 種名の重複
- シノニムの間違い
- スペルのミス  
などなど



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

# dct\_validate() が分類学的データの エラーを探し出してくれる

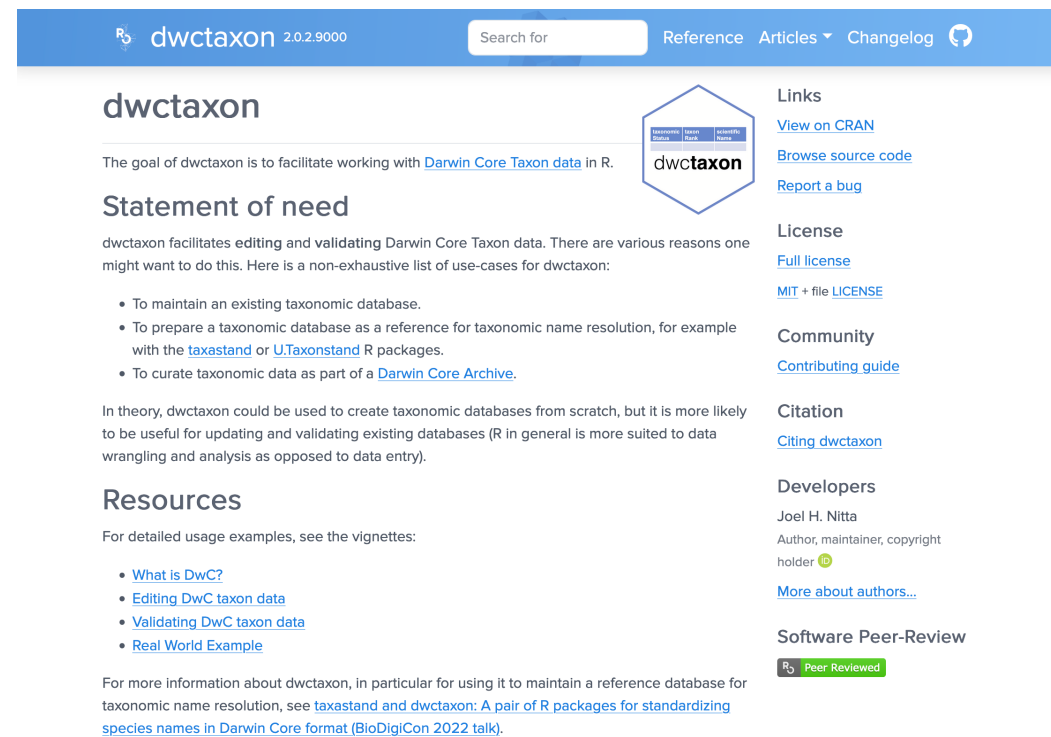
```
1 # Try to add a name that already exists
2 filmies |>
3   dct_add_row(scientificName = "Trichomanes crassum Copel.") |>
4   dct_validate()
```

```
Error: check_sci_name failed
  scientificName detected with duplicated value
  Bad scientificName: Trichomanes crassum Copel., Trichomanes crassum Copel.
```



# サマリー

- dwctaxonはDwC形式の**分類学的データを管理するRパッケージ**である
- 自動的にデータの**編集と検証**を行う
- 分類学的データのデータベースを管理している方や種名を扱っている方におすすめ
- インストールと使い方の詳細：  
<https://github.com/ropensci/dwctaxon>



The screenshot shows the CRAN page for the R package 'dwctaxon' version 2.0.2.9000. The page includes a search bar, navigation links for Reference, Articles, and Changelog, and a logo for dwctaxon. The main content area features the package name, a description of its goal to facilitate working with Darwin Core Taxon data, a 'Statement of need' section with three bullet points, a 'Resources' section with four links, and a 'Software Peer-Review' badge indicating it is 'Peer Reviewed'. A right-hand sidebar contains links for View on CRAN, Browse source code, Report a bug, License (Full license), MIT + file LICENSE, Community (Contributing guide), Citation (Citing dwctaxon), and Developers (Joel H. Nitta).

ご清聴ありがとうございます！

# 謝辞

- 日本学術振興会
- 東京大学岩崎研のメンバー