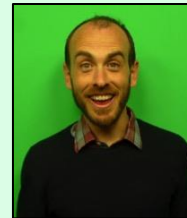
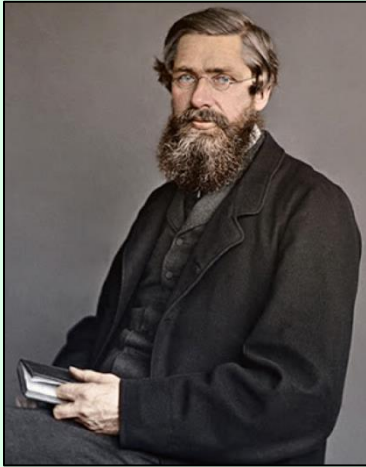


# *Wallace*: A flexible platform for reproducible modeling of species niches and distributions built for community expansion

Jamie M. Kass, Bruno Vilela, Matthew Aiello-Lammens, Robert Muscarella, and Robert P. Anderson





# Wallace

The screenshot shows the Wallace web application interface. The top navigation bar includes tabs for 'Wallace', 'Intro', '1 Occ Data', '2 Process Occs', '3 Env Data', '4 Process Envs', '5 Partition Occs', '6 Model', '7 Visualize', '8 Project', and 'Session Code'. The main content area is divided into several sections:

- Obtain Occurrence Data:** Includes 'Query Database' and 'Userspecified Occurrences' options.
- Module: Query Database:** A section for searching the database, with a sub-section 'via spocc package: Interface to Species Occurrence Data Sources'.
- Choose Database:** Options for 'GRIP', 'VertNet', and 'BISON'.
- Enter scientific name (format: Genus species):** A search input field containing 'Tremarctos ornatus'.
- Search Database:** A button to execute the search.
- Maximum number of occurrences:** A slider set to 1000.
- Download DB Occurrences:** A button to download the results.
- spocc references:** A list of developers and references.

On the right side, there is a map of South America showing the distribution of 'Tremarctos ornatus' with red dots. The map includes labels for 'COLOMBIA', 'PERU', 'AMAZON BASIN', and 'BRAZIL'. A search status box at the top right indicates: 'Searching gdf...', 'Total gdf records for Tremarctos ornatus returned [ 75 ] out of [ 209 ] total limit [ 75 ]. Records without coordinates removed [ 53 ]. Duplicated records removed [ 0 ]. Remaining records [ 22 ].' There are also buttons for 'Map', 'Occs Tbl', 'Results', 'Component Guidance', and 'Module Guidance'. A 'Change Base Map' dropdown is set to 'ESRI Topo'.

## What is *Wallace*? And Why?

# Applied biodiversity informatics

Anderson (2012)

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

## **Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions**

Robert P. Anderson<sup>1,2,3,4</sup>

*Agenda: Making data and modeling ready to address critical environmental issues of the 21st century*

# Agenda: applied biodiversity informatics

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

## **Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions**

Robert P. Anderson<sup>1,2,3,4</sup>

1. High-quality **data**, ready to be accessed when the particular problem presents itself
2. **Software** that achieves an appropriate balance between automation and supervision
3. **Scientists** capable of building, applying, and appraising high-quality models

# Agenda: applied biodiversity informatics

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

## **Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions**

Robert P. Anderson<sup>1,2,3,4</sup>

1. High-quality **data**, ready to be accessed when the particular problem presents itself
2. **Software** that achieves an appropriate balance between automation and supervision
3. **Scientists** capable of building, applying, and appraising high-quality models

# Agenda: applied biodiversity informatics

ANNALS OF THE NEW YORK ACADEMY OF SCIENCES

Issue: *Blavatnik Awards for Young Scientists*

## **Harnessing the world's biodiversity data: promise and peril in ecological niche modeling of species distributions**

Robert P. Anderson<sup>1,2,3,4</sup>

1. High-quality **data**, ready to be accessed when the particular problem presents itself

2. **Software** that achieves an appropriate balance between automation and supervision

3. **Scientists** capable of building, applying, and appraising high-quality models

# Agenda: applied biodiversity informatics

1. High-quality **data**, *ready to be accessed when the particular problem presents itself*

Continued large-scale, across-the-board databasing and georeferencing initiatives

High priority taxonomic projects realized in parallel

# Agenda: applied biodiversity informatics

2. **Software** that achieves an appropriate balance between automation and supervision

- *automates* repetitive aspects
- *forces* user to make critical biological and conceptual decisions
- *general* with respect to the algorithm/s used



# Agenda: applied biodiversity informatics

2. **Software** that achieves an appropriate balance between automation and supervision

- *automates* repetitive aspects
- *forces* user to make critical biological and conceptual decisions
- *general* with respect to the algorithm/s used

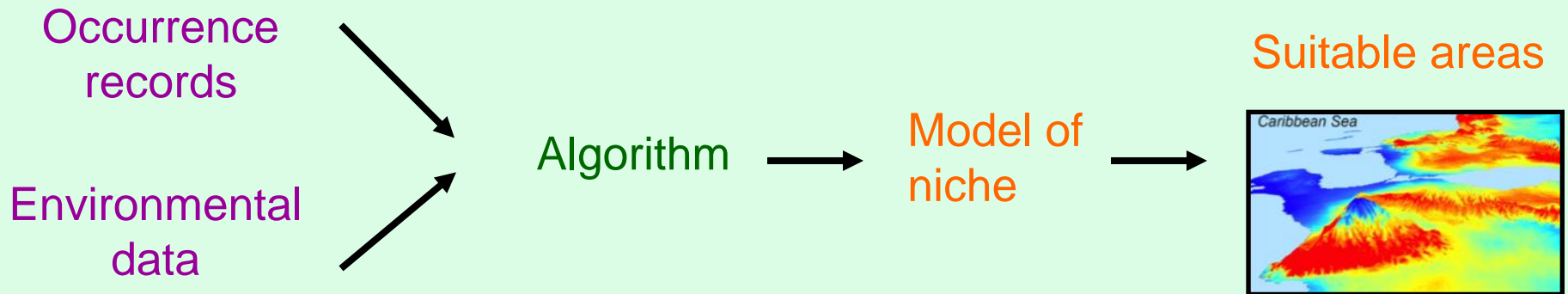
*“umbrella tools that can be employed with any particular modeling algorithm”*

# Agenda: applied biodiversity informatics

3. **Scientists** capable of building, applying, and appraising high-quality models

- clear literature
- workshops
- graduate courses

# Correlative modeling of species niches and distributions



## Anderson (2015)

11

BIOGEOGRAFÍA 8

LEAD ARTICLE

### **MODELING NICHEs AND DISTRIBUTIONS: IT'S NOT JUST "CLICK, CLICK, CLICK"**

**Robert P. Anderson**

#### **INTRODUCTION**

I offer this short essay with the idea of emphasizing that, despite being a relatively new field still under development, ecological niche modeling (and its applications for studies of

most important tools in biogeography and related fields (Guisan & Thuiller 2005, Wiens & Graham 2005, Elith & Leathwick 2009, Franklin 2010, Peterson et al. 2011), but I feel that it is not clear to many researchers just how tight the

# Recent R packages: spThin and ENMeval


000010000  
000101000  
0001000100  
0010000100  
0100000100  
0100000010  
1000000001  
Software notes

**Ecography** 38: 001–005, 2015  
doi: 10.1111/ecog.01132  
© 2015 The Authors. Ecography © 2015 Nordic Society Oikos  
Subject Editor: Thiago Rangel. Editor-in-Chief: Miguel Araújo. Accepted 18 November 2014

## spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models

Matthew E. Aiello-Lammens, Robert A. Boria, Aleksandar Radosavljevic, Bruno Vilela and Robert P. Anderson

*M. E. Aiello-Lammens (matt.lammens@gmail.com), Dept of Ecology and Evolutionary Biology, Univ. of Connecticut, Storrs, CT 06269, USA, and Dept of Ecology and Evolution, Stony Brook Univ., Stony Brook, NY 11794, USA. – R. A. Boria, A. Radosavljevic and R. P. Anderson, Dept of Biology, City College of the City Univ. of New York, New York, NY 10031, USA. AR present address: Plant Biology and Conservation, Northwestern Univ., Evanston, IL 60208, USA, and Dept of Plant Science, Chicago Botanic Garden, Glencoe, IL 60022, USA, and Dept of Botany, National Museum of Natural History, Smithsonian Inst., Washington, DC 20560, USA. RPA also at: Graduate Center of the City Univ. of New York, New York, NY 10016, USA, and Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, New York, NY 10024, USA. – B. Vilela, Depto de Ecologia, Inst. de Ciências Biológicas, Univ. Federal de Goiás, Goiânia, Goiás, Brazil, and Depto de Ciencias de la Vida, Univ. de Alcalá, ES-28805 Alcalá de Henares, Madrid, Spain.*

**Methods in Ecology and Evolution** 

*Methods in Ecology and Evolution* 2014, 5, 1198–1205  
doi: 10.1111/2041-210X.12261

**APPLICATION**

## ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models

Robert Muscarella<sup>1\*</sup>, Peter J. Galante<sup>2</sup>, Mariano Soley-Guardia<sup>2,3</sup>, Robert A. Boria<sup>2</sup>, Jamie M. Kass<sup>2,3</sup>, Maria Uriarte<sup>1</sup> and Robert P. Anderson<sup>2,3,4</sup>

<sup>1</sup>Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Ave., New York, NY 10027, USA; <sup>2</sup>Department of Biology, City College of the City University of New York, 160 Convent Ave., New York, NY 10031, USA; <sup>3</sup>Graduate Center of the City University of New York, 365 5th Ave., New York, NY 10016, USA; and <sup>4</sup>Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West & 79th Street, New York, NY 10024, USA



Matthew Aiello-Lammens



Robert Muscarella

Robert Anderson, Robert Boria, Peter Galante, Jamie Kass,  
Aleksandar Radosavljevic, Mariano Soley-Guardia, Maria Uriarte, Bruno Vilela



Software notes

Ecography 38: 001–005, 2015

doi: 10.1111/ecog.01132

© 2015 The Authors. Ecography © 2015 Nordic Society Oikos

Subject Editor: Thiago Rangel. Editor-in-Chief: Miguel Aráujo. Accepted 18 November 2014

## spThin: an R package for spatial thinning of species occurrence records for use in ecological niche models

Matthias  
Roberts

*M. E. A.  
and D.  
Deps of  
Northwest  
Botany,  
of New  
NY 100  
Ciencia*

## Methods in Ecology and Evolution



Methods in Ecology and Evolution 2014, 5, 1198–1205

doi: 10.1111/2041-210X.12261

### APPLICATION

## ENMeval: An R package for conducting spatially independent evaluations and estimating optimal model complexity for MAXENT ecological niche models

Robert Muscarella<sup>1\*</sup>, Peter J. Galante<sup>2</sup>, Mariano Soley-Guardia<sup>2,3</sup>, Robert A. Boria<sup>2</sup>,  
Jamie M. Kass<sup>2,3</sup>, Maria Uriarte<sup>1</sup> and Robert P. Anderson<sup>2,3,4</sup>

<sup>1</sup>Department of Ecology, Evolution and Environmental Biology, Columbia University, 1200 Amsterdam Ave., New York, NY 10027, USA; <sup>2</sup>Department of Biology, City College of the City University of New York, 160 Convent Ave., New York, NY 10031, USA; <sup>3</sup>Graduate Center of the City University of New York, 365 5th Ave., New York, NY 10016, USA; and <sup>4</sup>Division of Vertebrate Zoology (Mammalogy), American Museum of Natural History, Central Park West & 79th Street, New York, NY 10024, USA

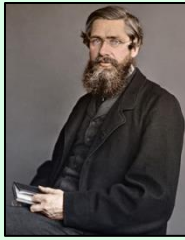






*Creativity,*  
discernment,  
*“discussion”*  
and FUN





# Wallace

1

Wallace

Intro

1 Occ Data

2 Process Occs

3 Env Data

4 Process Envs

5 Partition Occs

6 Model

7 Visualize

8 Project

Session Code

Obtain Occurrence Data

Modules Available:

- Query Database
- User-specified Occurrences

## Module: Query Database

via **spocc** package: Interface to Species Occurrence Data Sources

Choose Database:

- GBIF
- VertNet
- BISON

Enter scientific name (format: Genus species)

Tremarctos ornatus

Search Database

Maximum number of occurrences:



Download DB Occurrences

spocc references

Developers: Scott Chamberlain, Karthik Ram, Ted Hart

[CRAN](#) | [documentation](#)

... Searching gbif ...

\* Total gbif records for Tremarctos ornatus returned [ 75 ] out of [ 209 ] total (limit 75 ). Records without coordinates removed [ 53 ]. Duplicated records removed [ 0 ]. Remaining records [ 22 ].

Map Occs Tbl Results Component Guidance Module Guidance



Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, TomTom, Intermap, iPC, USGS, FAO, NPS, NRCAN, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, METI, Esri China (Hong Kong), and the GIS User Community

2

2a

2b

3

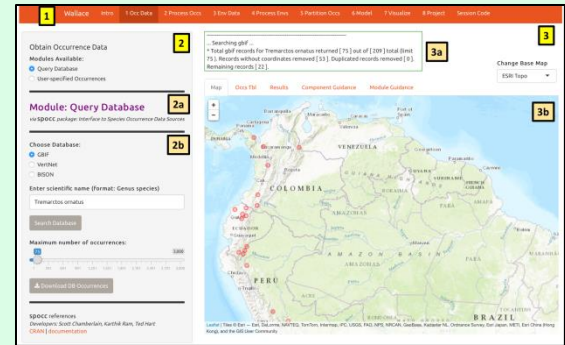
3a

3b

Change Base Map

ESRI Topo

# What is *Wallace* beta 0.6?



Point-and-click (GUI) application that includes most steps of a niche/distribution modeling workflow

- Harnesses R packages and gives credit to their authors

# Qualities of *Wallace*

*Wallace's* code is free  
and **OPEN**

(& users can download  
data from online  
databases.)

1 Wallace Intro 1 Occ Data 2 Process Occs

Obtain Occurrence Data 2

Modules Available:

- Query Database
- User-specified Occurrences

---

Module: Query Database 2a

via **spocc** package: Interface to Species Occurrence Data Sources

---

Choose Database: 2b

- GBIF
- VertNet
- BISON

Enter scientific name (format: Genus species)

Tremarctos ornatus

Search Database

Maximum number of occurrences:

75 3,000

1 301 601 901 1,201 1,501 1,801 2,101 2,401 2,701 3,000

Download DB Occurrences

---

spocc references

Developers: Scott Chamberlain, Karthik Ram, Ted Hart

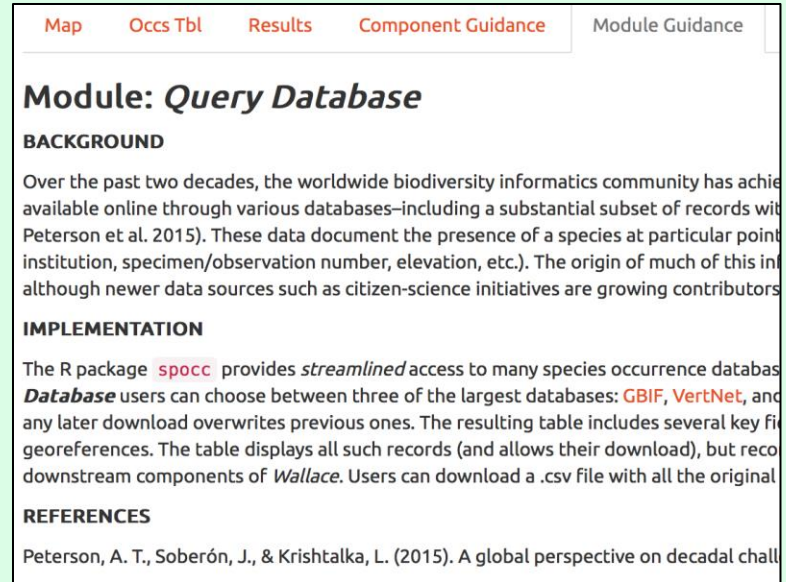
[CRAN](#) | [documentation](#)

# Qualities of *Wallace*

*Wallace* provides

***GUIDANCE***

that addresses  
conceptual and  
methodological  
issues.



The screenshot shows a web interface with a navigation bar at the top containing the following tabs: [Map](#), [Occs Tbl](#), [Results](#), [Component Guidance](#) (which is the active tab), and [Module Guidance](#). Below the navigation bar, the page title is **Module: *Query Database***. The content is organized into three sections: **BACKGROUND**, **IMPLEMENTATION**, and **REFERENCES**. The **BACKGROUND** section discusses the worldwide biodiversity informatics community and the availability of species occurrence data. The **IMPLEMENTATION** section describes the R package `spocc` and its access to databases like GBIF and VertNet. The **REFERENCES** section lists a paper by Peterson et al. (2015).

[Map](#) [Occs Tbl](#) [Results](#) [Component Guidance](#) [Module Guidance](#)

## Module: *Query Database*

### BACKGROUND

Over the past two decades, the worldwide biodiversity informatics community has achieved significant progress in making species occurrence data available online through various databases—including a substantial subset of records with georeferences (Peterson et al. 2015). These data document the presence of a species at particular points in time and space (e.g., institution, specimen/observation number, elevation, etc.). The origin of much of this information is from museum collections, although newer data sources such as citizen-science initiatives are growing contributors.

### IMPLEMENTATION

The R package `spocc` provides *streamlined* access to many species occurrence databases. ***Database*** users can choose between three of the largest databases: **GBIF**, **VertNet**, and **VertNet**. Any later download overwrites previous ones. The resulting table includes several key fields, including species names and georeferences. The table displays all such records (and allows their download), but records are not yet integrated into the downstream components of *Wallace*. Users can download a .csv file with all the original data.

### REFERENCES

Peterson, A. T., Soberón, J., & Krishtalka, L. (2015). A global perspective on decadal challenges in biodiversity informatics.

# Qualities of *Wallace*

*Wallace* is **FLEXIBLE**

by providing multiple options, and allowing user inputs and downloads for most components.

### Module: Select Study Region

via `sp` and `rgeos` packages: *Title Classes and Methods for Spatial Data | Interface to Geometry Engine*  
- Open Source (GEOS)

---

**Background Extents:**

Bounding box

Minimum convex polygon

User-specified polygon

**Study region buffer distance (degree)**

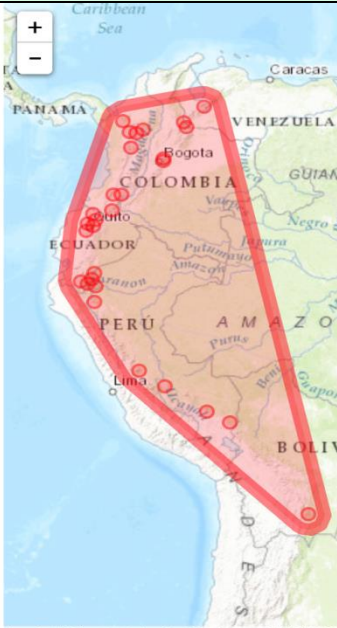
1

Clip Env Data by Polygon

**Select File Type**

GRD

Download Clipped Env Data



Leaflet | Tiles © Esri — Esri, DeLorme, NAVTEQ, GeoBase, Kadaster NL, Ordnance Survey, Esri Japan, Swisstopo, Mapbox, Mapbox, OpenStreetMap, SRTM Source: Esri, i-cubed, USDA, USGS, AEX, GeoEye

# Qualities of *Wallace*

*Wallace* features

***INTERACTIVE***

maps, tables, and  
graphs to explore  
data and model  
predictions.



# Qualities of *Wallace*

*Wallace* is

***REPRODUCIBLE***

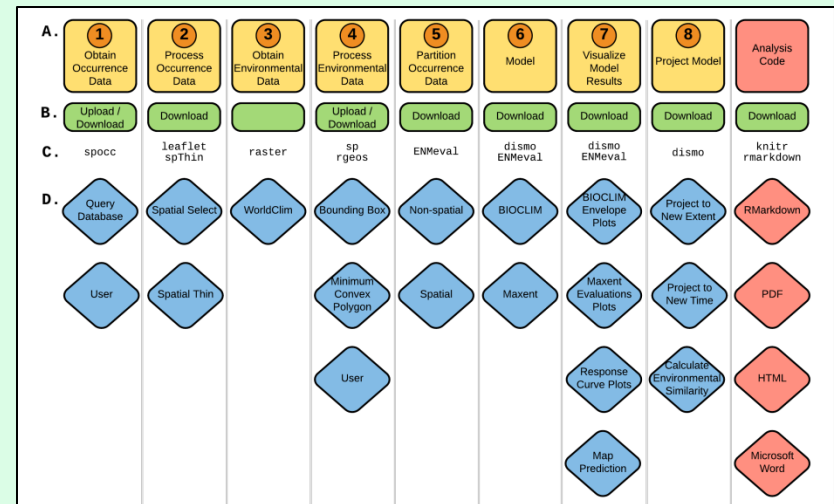
by providing  
executable code for  
documenting and  
rerunning the  
analysis.

```
1 Please find below the R code history from your "Wallace" session. You can reproduce your session
2 file in RStudio (for more information see <http://rmarkdown.rstudio.com>).
3
4 ## Package installation
5 Wallace uses the following R packages that must be installed and loaded before starting.
6
7 ```{r}
8 library(devtools)
9 library(spocc)
10 library(maptools)
11 library(spThin)
12 library(dismo)
13 library(rgeos)
14 library(repms)
15 library(maps)
16 library(ENMeval)
17 ```
18
19 Wallace also includes several functions developed to help integrate different packages and some of
20 this reason, it is necessary to load the file 'functions.R', which can be found on Wallace's GitHub
21 (<https://github.com/wallaceEcoMod/wallace>). Download the file, place it in your working directory
22 and then load it:
23
24 ```{r}
25 source(file.path("~/Users/musabali/Documents/github/wallace", "functions.R"))
26 ```
27
28 Record of analysis for "Tremarctos ornatus".
29 -----
30 ## Obtain Occurrence Data
31
32 The search for occurrences was limited to 81 records. Obtain occurrence records of the selected species.
33
34 ```{r}
35 results <- occ(query = "Tremarctos ornatus", from = "gbif", limit = 81, has_coords = TRUE)
36 results.data <- results[["gbif"]][,data[["formatSpName("Tremarctos ornatus")"]]
37 occs <- remDups(results.data) # remove rows with duplicate coordinates
38 ```
39
40 ## Process Occurrence Data
41
42 You chose 42 of 51 total occurrence localities via polygon selection to keep in the analysis.
43
44 ```{r}
45 occs <- occs[c(1, 2, 3, 5, 6, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 20, 21, 22, 24, 25, 26, 27,
46 40, 41, 42, 43, 44, 45, 46, 47, 48, 49, 50, 51), ] # subset occs by selected rows
```

# Qualities of *Wallace*

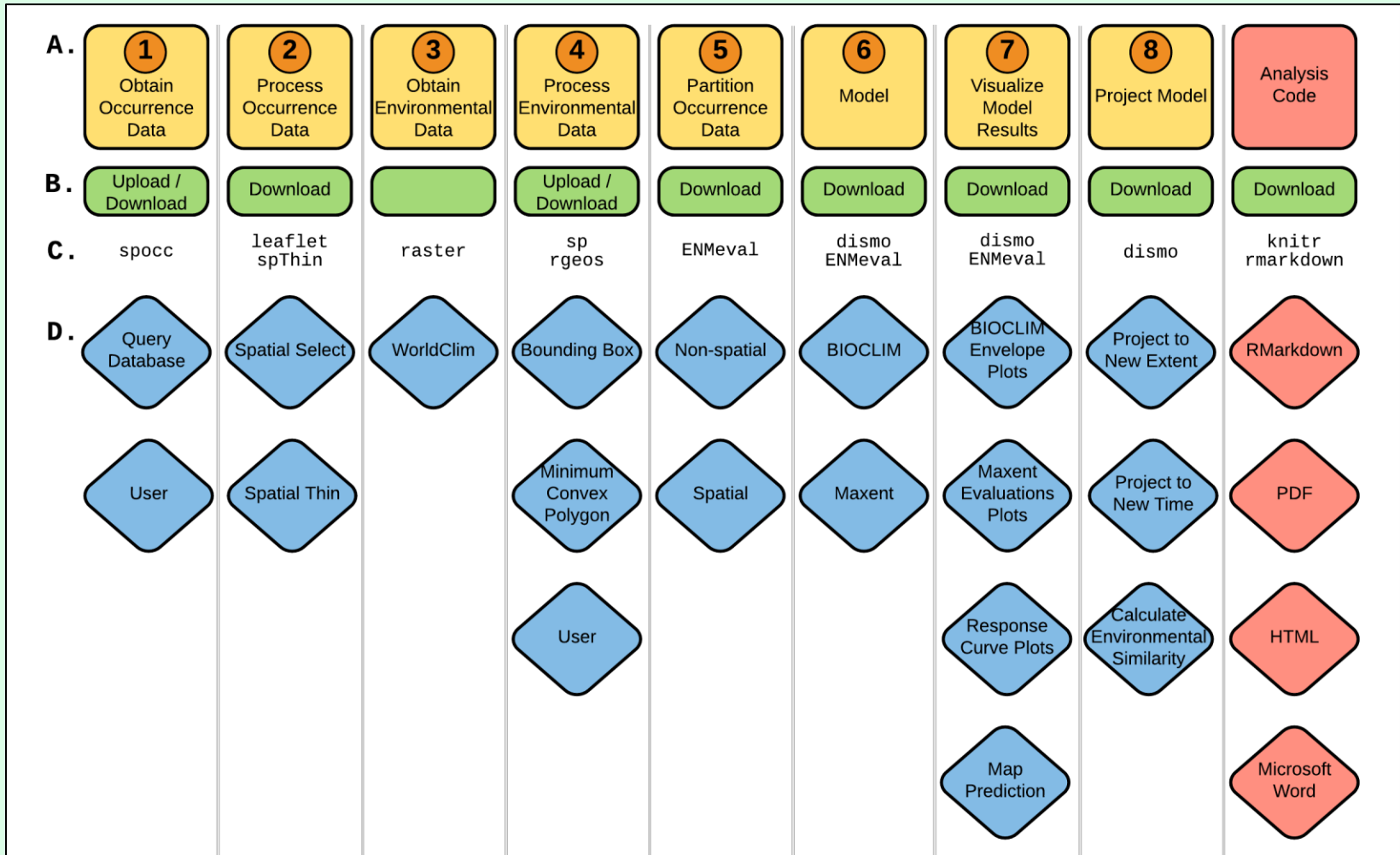
*Wallace* is **EXPANDABLE!**

by featuring different methodological options (modules) that researchers can contribute to advance functionality.

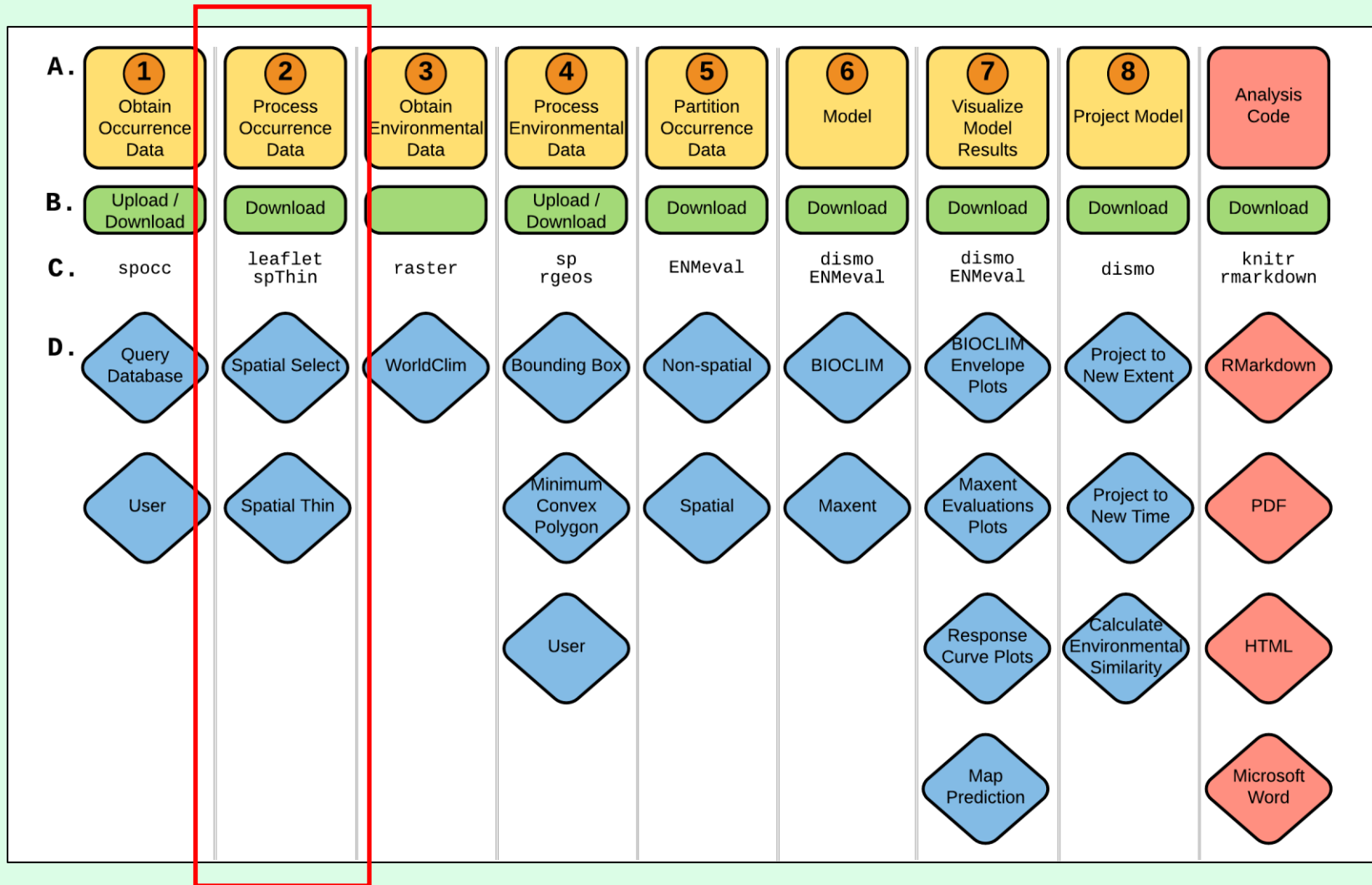




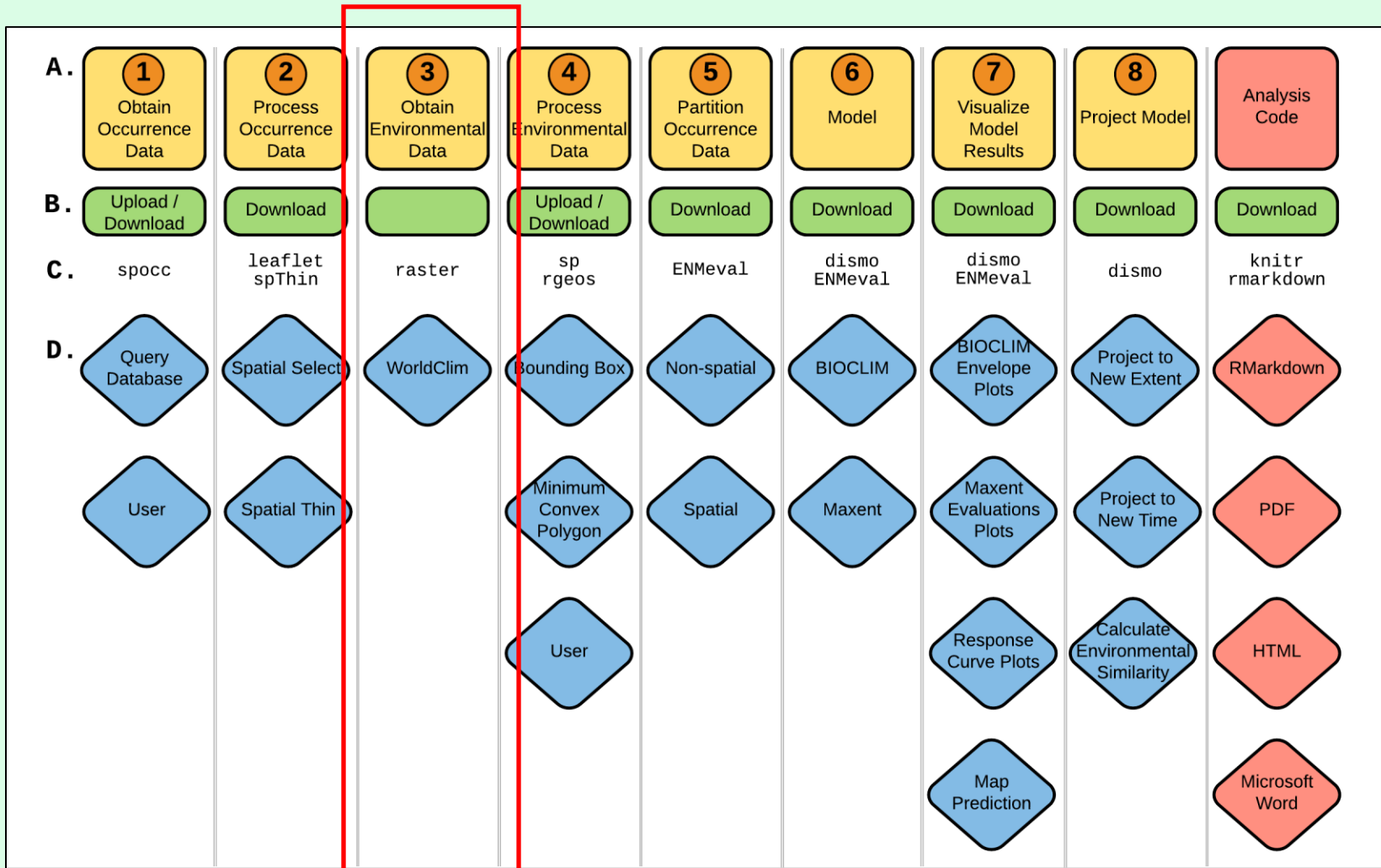
# Wallace is expandable



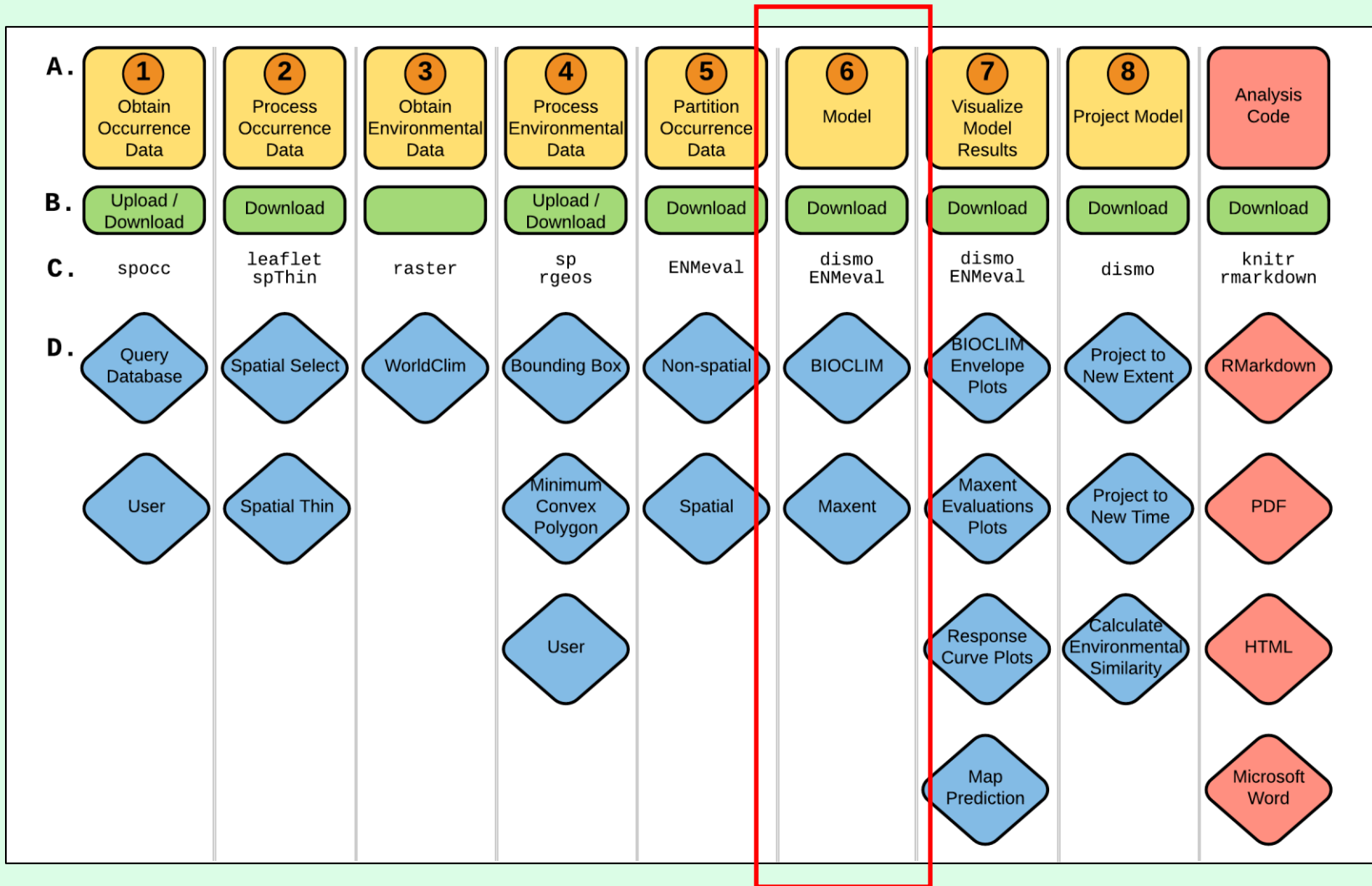
# Wallace is expandable



# Wallace is expandable



# Wallace is expandable

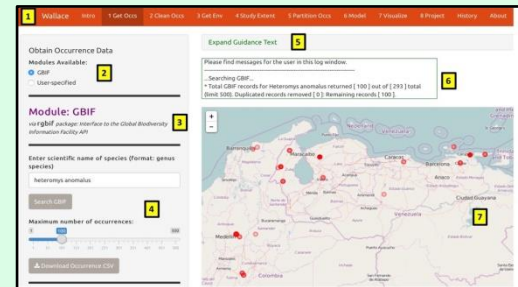


# Future directions

1. First *full release soon*  
(software note in review)

2. Ongoing NSF funding to respond to immediate user needs (*Google group and Wallace e-mail*)

3. Seeking more funding, *with Cory Merow* (to work with external partners to add new modules)



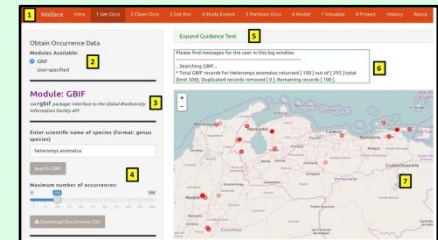
NSF DBI-1650241

# THANK YOU

... please help us make *Wallace* better ... and grow



DEB-1119915 and DBI-1650241



<http://wallaceecomod.github.io>  
<http://www.andersonlab.cuny.cuny.edu>

