

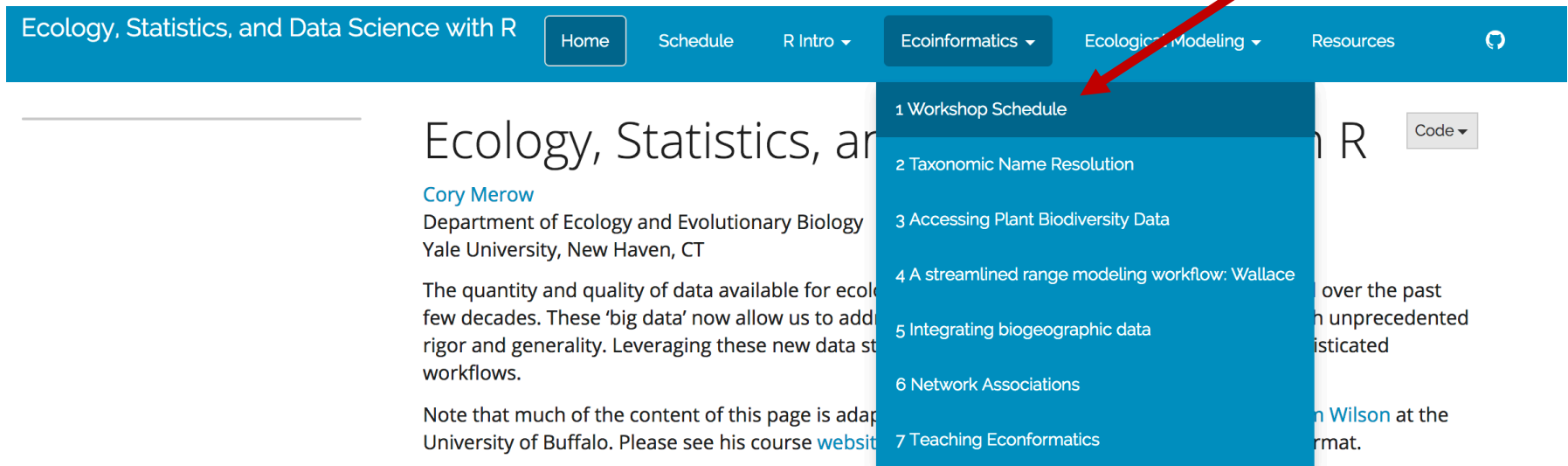
Integrating and Cleaning Biodiversity Data

A bunch of ecoinformaticians (see next)

9 Jan 2017 IBS Meeting

Install some packages while I'm talking...

cmerow.github.io/RDataScience/



The screenshot shows the website 'Ecology, Statistics, and Data Science with R'. The navigation bar includes links for Home, Schedule, R Intro, Ecoinformatics, Ecological Modeling, and Resources. A red arrow points to the 'Ecoinformatics' dropdown menu, which is open and displays a list of seven items: 1 Workshop Schedule, 2 Taxonomic Name Resolution, 3 Accessing Plant Biodiversity Data, 4 A streamlined range modeling workflow: Wallace, 5 Integrating biogeographic data, 6 Network Associations, and 7 Teaching Ecoinformatics. The main content area features a header 'Ecology, Statistics, and Data Science with R' by Cory Merow, with a 'Code' button. The text below the header discusses the quantity and quality of data available for ecology and the use of 'big data' in workflows.

Ecology, Statistics, and Data Science with R

Cory Merow
Department of Ecology and Evolutionary Biology
Yale University, New Haven, CT

The quantity and quality of data available for ecology has increased significantly over the past few decades. These 'big data' now allow us to add rigor and generality. Leveraging these new data streams into streamlined workflows.

Note that much of the content of this page is adapted from the course taught by John Wilson at the University of Buffalo. Please see his course [website](#).

1 Workshop Schedule

2 Taxonomic Name Resolution

3 Accessing Plant Biodiversity Data

4 A streamlined range modeling workflow: Wallace

5 Integrating biogeographic data

6 Network Associations

7 Teaching Ecoinformatics

Code



Matt Aiello-Lammens



Rob Anderson



Brad Boyle



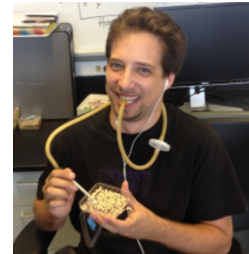
Brian Enquist



Jamie Kass



Drew Kerkhoff



Brian Maitner



Brian McGill



Cory Merow



Naia Morueta-Holme

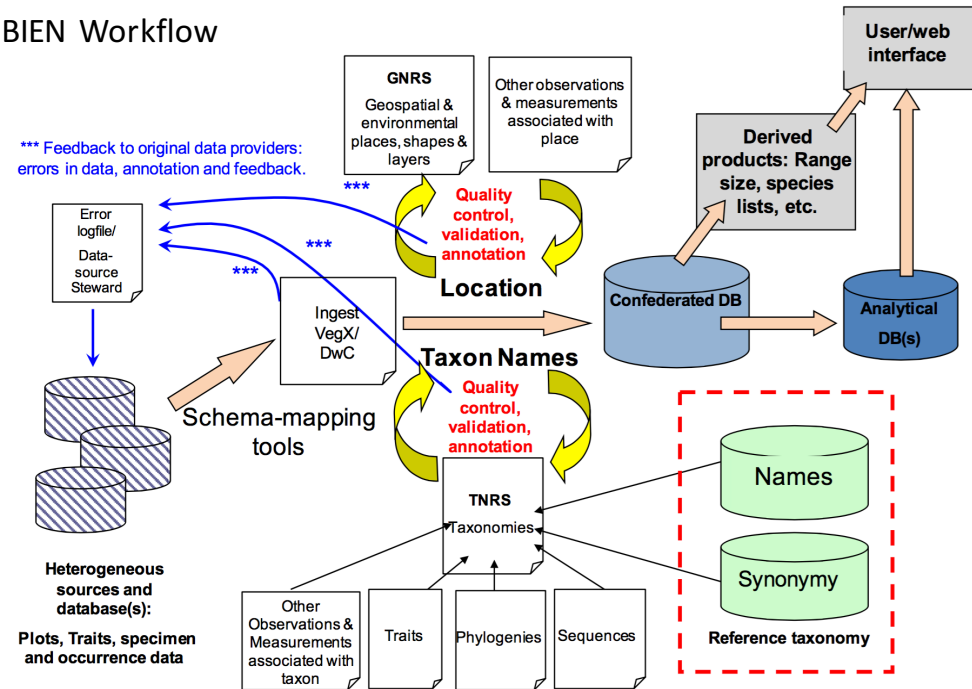


Jens Christian Svenning

What is informatics

- Informatics – getting info from data
 - **Computation**

BIEN Workflow



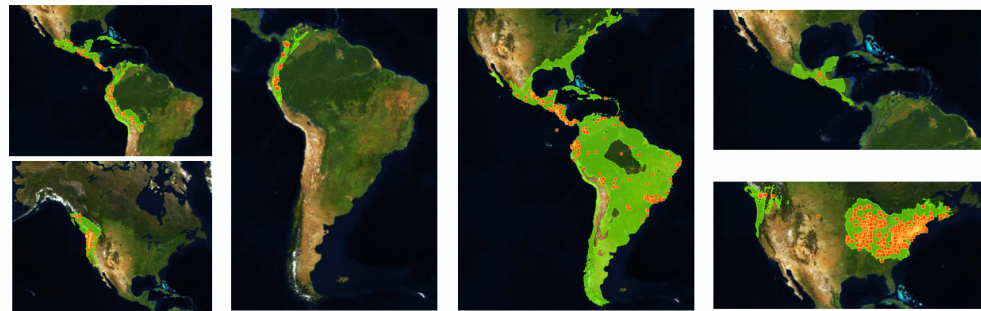
What is informatics

- Informatics – getting info from data
 - Computation
 - **Large data sets**

BEN Botanical Information and Ecology Network



~200,000 species: Occurrence, Communities, Traits
All New World Plants



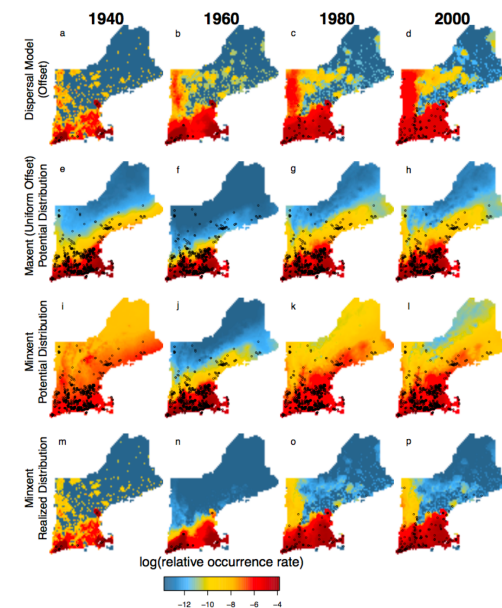
What is informatics

- Informatics – getting info from data
 - Computation
 - Large data sets
 - **Statistics**

New Models - Minxent

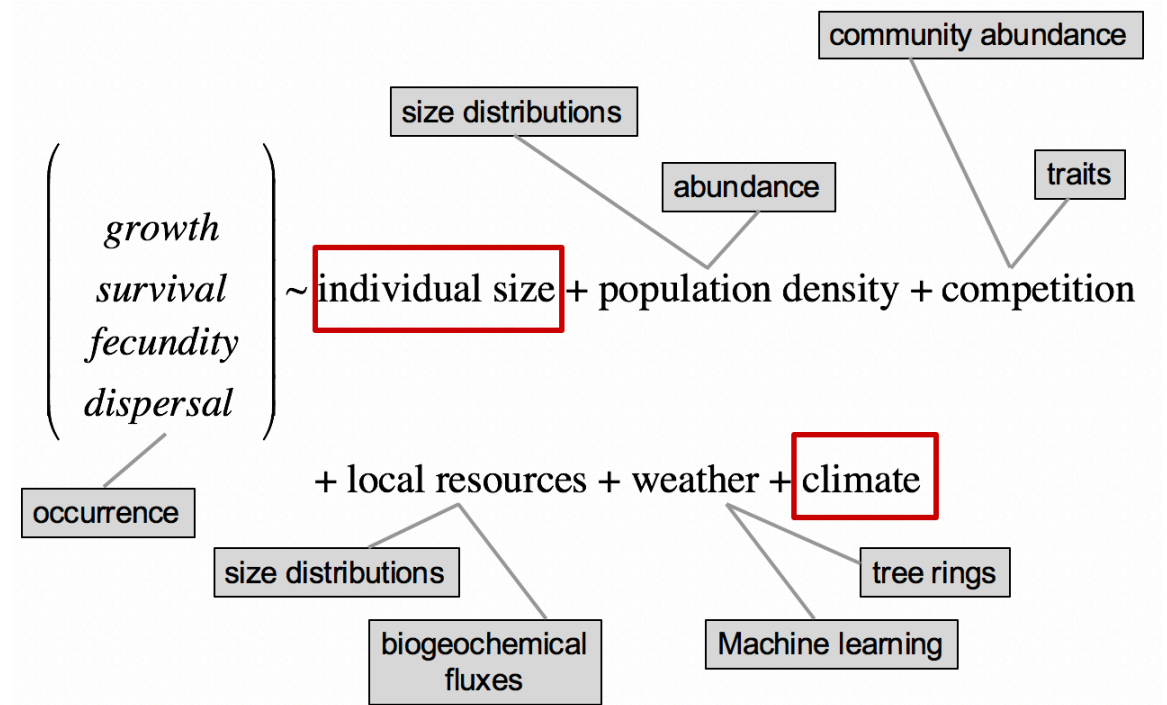
- Sampling Bias
- Native range info
 - Realized vs
 - Potential Distributions
- Expert maps
- Phylogenetic constraints

Merow et al. GEB 2016



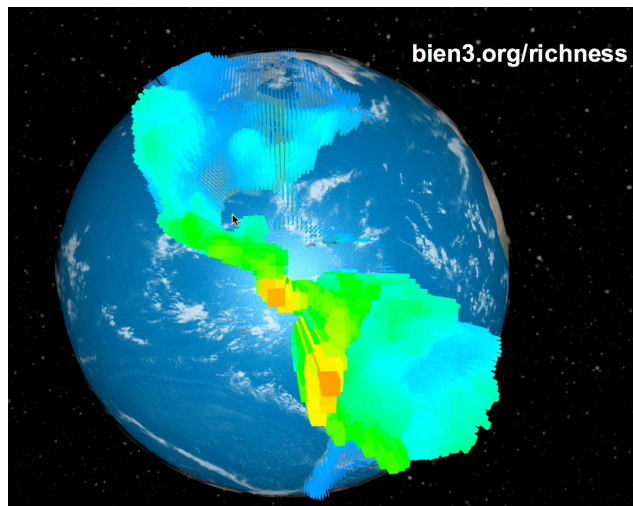
A case for informatics

- **Combine data sets**

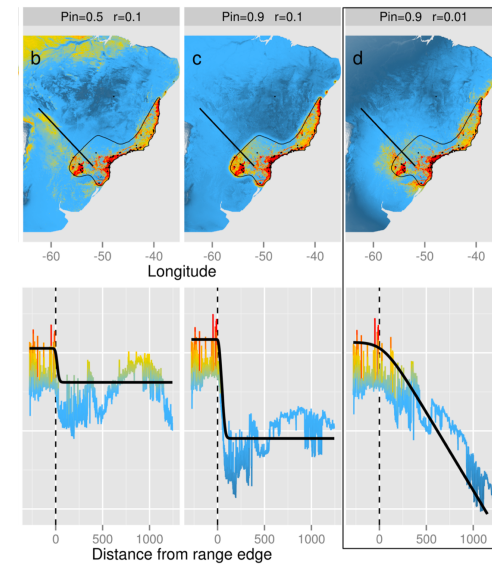


A case for informatics

- Combine data sets
- **Explore data sets**



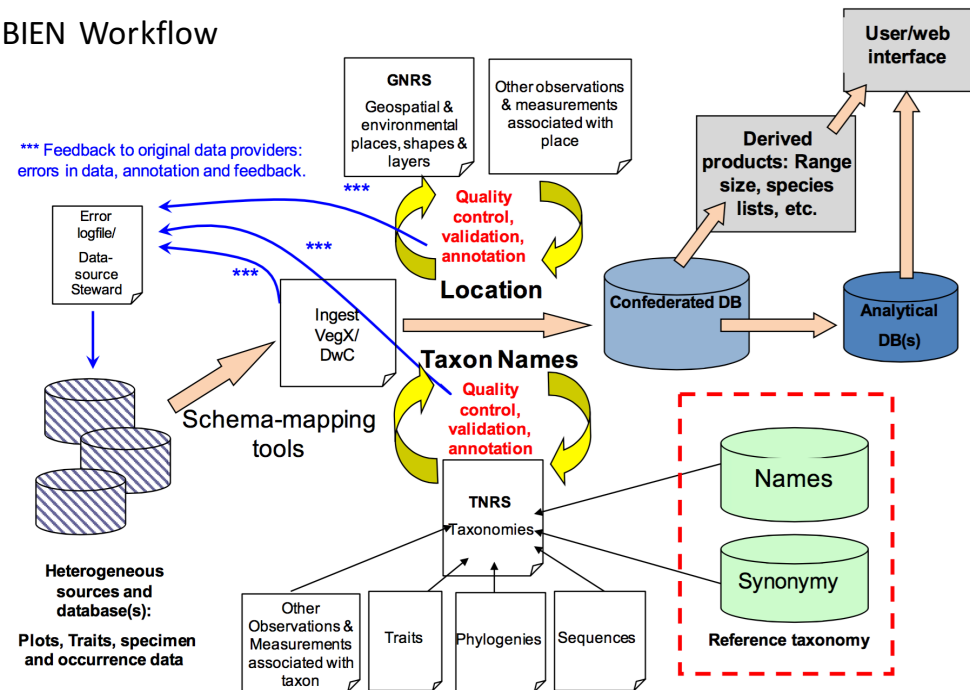
Association network



A case for informatics

- Combine data sets
- Explore data sets
- **Get our data organized**

BIEN Workflow



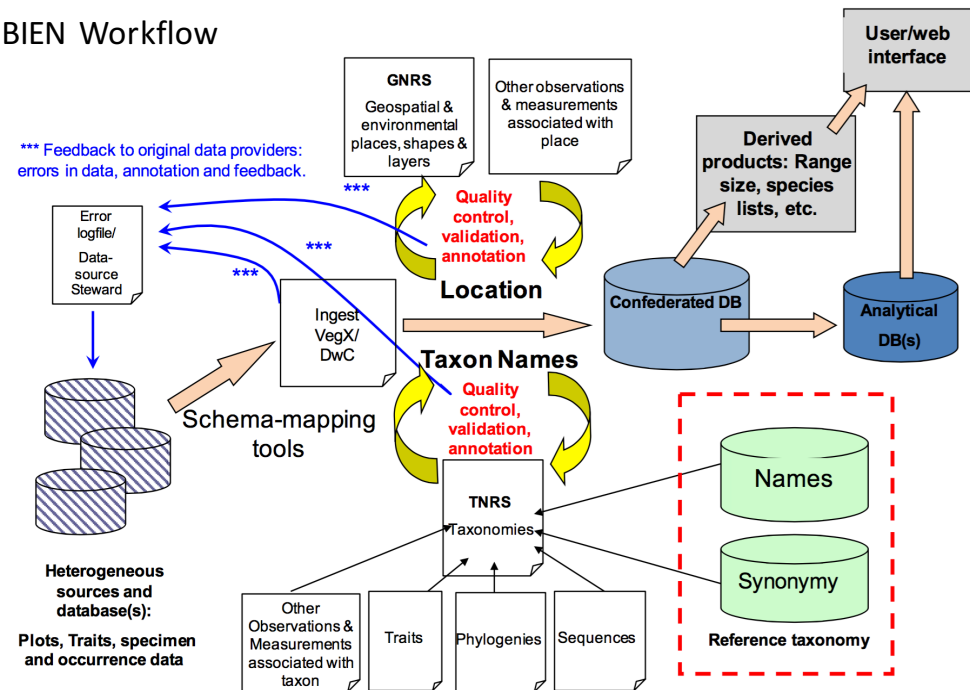
A case for informatics

- Combine data sets
- Explore data sets
- **Get our data organized**

RBien or BRI?

```
BIEN_occurrence_species()  
BIEN_ranges_load_species()  
BIEN_plot_dataset()
```

BIEN Workflow



A case for informatics

- Combine data sets
- Explore data sets
- **Get our data organized**

RBien or BRI?

```
BIEN_occurrence_species()  
BIEN_ranges_load_species()  
BIEN_plot_dataset()
```


WALLACE beta v0.2: An R-based Modular Web App to Harness Biodiversity Data for Spatial Modeling
Developers: Jamie M. Kass, Matthew Aiello-Lammens, Bruno Vilela, Robert Muscarella, Robert P. Anderson

Introduction | **1) Obtain Occ Data** | 2) Process Occ Data | 3) Obtain Env Data
4) Process Env Data | 5) Partition Occ Data | 6) Build Niche Model | 7) Visualize Results

[Download Code History](#) | [About](#)

LOG

The output of this step is a CSV file with rows of localities, and columns containing species, longitude, and latitude (as well as any other fields provided by GBIF or the user).



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^{1,2,3,4}

*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^{1,2,3,4}

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

THANK YOU



NSF DEB-1119915 and DBI-1650241



<http://www.andersonlab.ccnycuny.edu/>



cmerow.github.io

