

« Biological and environmental data »



REBIOMA

Rebioma@rebioma.net

- Data Sources
- Data Quality / Georeferencing
- Georeferencing
- Management: Rebioma process

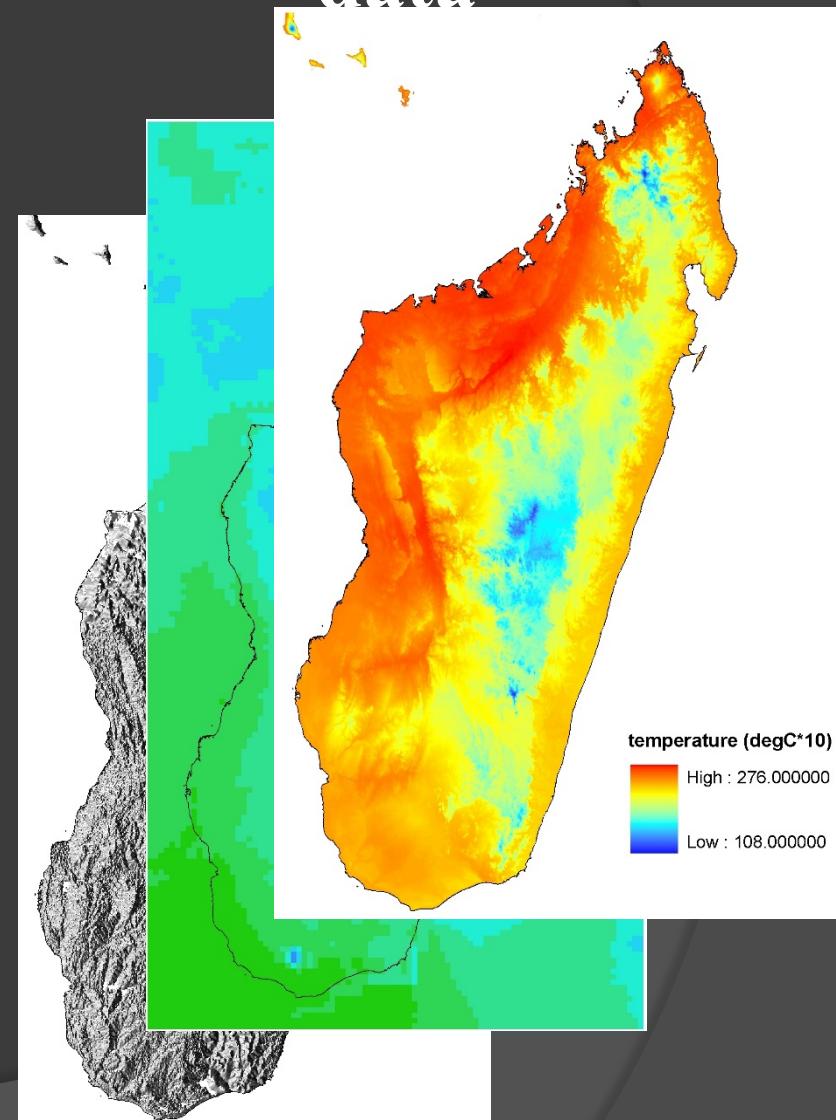
Species' distribution data



Uroplatus sp.
(leaf-tailed
gecko)



Environmental data



Biodiversity Data

- ◎ **Primary biodiversity data :**

Place a particular taxon at a particular place at a particular time with appropriate documentation

- ◎ **Secondary biodiversity data :**

Information synthesized (often from primary data) to summarize data



Research-grade Data?

- Data are often « dumbed down » ... i.e., simplified, organized, synthesized, or summarized
- Such data are easy to interpret and understand, but generally do not suffice as a basis for detailed analysis
- Require research-grade primary data ... on which a scientific publication or detailed report could be based



Biodiversity data management



DarwinCore

What is the Darwin Core (DwC)?

- Standard for sharing information about species and biodiversity information
- Based on taxa, their occurrence in nature as documented by observations, specimens, and samples, and related information
- Includes a glossary of terms that provide reference definitions, examples, and commentaries, including how terms
 - are managed
 - can be extended for new purposes
 - can be used
- Design Philosophy: minimize the barriers to adoption and to maximize reusability

(cf. <http://rs.tdwg.org/dwc/terms/index.htm>)

Format

→ **Minimum field required :**

- YearCollected: Année de collecte
- Genus: genre
- SpecificEpithet: espèce
- DecimalLatitude
- DecimalLongitude
- GeodeticDatum
- CoordinateUncertaintyInMeters,
- NomenclaturalCode.

Darwin Core: An Evolving Community-Developed Biodiversity Data Standard

John Wieczorek¹, David Bloom^{1*}, Robert Guralnick², Stan Blum³, Markus Döring⁴, Renato Giovanni⁵, Tim Robertson⁴, David Vieglais⁶

1 University of California, Berkeley, California, United States of America, **2** University of Colorado, Boulder, Colorado, United States of America, **3** California Academy of Sciences, San Francisco, California, United States of America, **4** Global Biodiversity Information Facility, Copenhagen, Denmark, **5** Centro de Referência em Informação Ambiental, Campinas, São Paulo, Brasil, **6** University of Kansas, Lawrence, Kansas, United States of America

Abstract

Biodiversity data derive from myriad sources stored in various formats on many distinct hardware and software platforms. An essential step towards understanding global patterns of biodiversity is to provide a standardized view of these heterogeneous data sources to improve interoperability. Fundamental to this advance are definitions of common terms. This paper describes the evolution and development of Darwin Core, a data standard for publishing and integrating biodiversity information. We focus on the categories of terms that define the standard, differences between simple and relational Darwin Core, how the standard has been implemented, and the community processes that are essential for maintenance and growth of the standard. We present case-study extensions of the Darwin Core into new research communities, including metagenomics and genetic resources. We close by showing how Darwin Core records are integrated to create new knowledge products documenting species distributions and changes due to environmental perturbations.

Citation: Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. doi:10.1371/journal.pone.0029715

Editor: Indra Neil Sarkar, University of Vermont, United States of America

Received September 19, 2011; **Accepted** December 1, 2011; **Published** January 6, 2012

Copyright: © 2012 Wieczorek et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: The U.S. National Science Foundation (<http://www.nsf.gov/>) funded three individual projects that contributed to Darwin Core: Species Analyst Project (NSF-9808739), ORNIS (DBI-0345448), and MaNIS (DBI-0108161). The Gordon and Betty Moore Foundation provided individual grants to the BioGeomancer Project and to TDWG, both of which had influence on the Darwin Core (<http://www.moore.org/>). Global Biodiversity Information Facility, GBIF, sponsored a workshop to prepare Darwin Core for the standards track, in February 2009 (<http://www.gbif.org/>). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: dbloom@vertnet.org

Clean & Enhance Data with Tools

Open Refine, desktop app

- from messy to marvelous
- <http://code.google.com/p/google-refine/>
- <http://openrefine.org/>
- remove leading / trailing white spaces
- standardize values
- using scripts for repetitive data manipulation tasks
- ...





A power tool for working with messy data.

Create Project

Open Project

Import Project



Version 2.5 [r2407]

Help

About

Create a project by importing data. What kinds of data files can I import?

TSV, CSV, *SV, Excel (.xls and .xlsx), JSON, XML, RDF as XML, and Google Data documents are all supported. Support for other formats can be added with Google Refine extensions.

Get data from

This Computer

[Web Addresses
\(URLs\)](#)

[Clipboard](#)

[Google Data](#)

Locate one or more files on your computer to upload:

 [Browse...](#)

Next »

Georeferencing



What is Georeferencing?

A numerical description of a place that can be mapped and that describes the spatial extent of a locality and its associated uncertainties as well as possible.



Method Comparison

point

easy to produce
no data quality

bounding-box

simple spatial queries
difficult quality assessment

point-radius

easy quality assessment
difficult spatial queries

shape

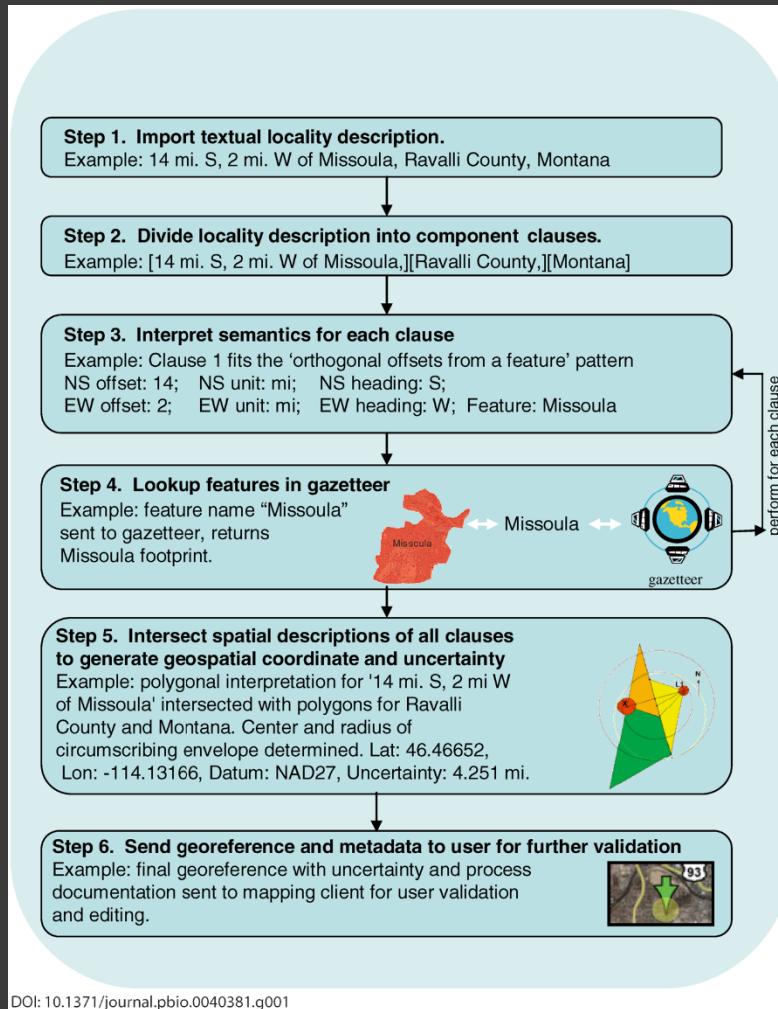
accurate representation
complex, uniform

probability

accurate representation
complex, non-uniform



BioGeomancer



DOI: 10.1371/journal.pbio.0040381.g001

Figure 1. Diagram showing how BioGeomancer converts a textual locality description into a geospatial description suitable for use in geographic information systems.

Why georeference?

- Map species locations
- Understand species ranges
- Enable spatial analyses
- Combine with other spatial data

Georeferencing Sources

- Specimen labels
- Field notes
- Literature
- Gazetteers
- Printed Maps
- Digital Maps

REBIOMA - *Data*: Online Data Portal

<http://data.rebioma.net>



ReBioMa Data Portal x Tsiky

data.rebioma.net/#tab=occ&view=Map&zoom=5¢er=-18.538148,56.2834473&map_type=terrain&left_tab=0&m_search=&m_page=1&page=1&type=all%20pos%20reviewed

Applications Bookmarks Funmoods - web sea Google Google Maps DISM ghost cast Windows7 MBR repai Disk partitioning Boîte de réception Ecoengine - spatial p Autres favoris

Version 1.0.r1580 | Sign in | Register | REBIOMA | Code site | Help | Acknowledgment | English ▾

REBIOMA

occurrences Species explorer Data dashboard

Search All Reliable Reviews for taxon Search Advanced Search Shapefile List CSV

Select: All None

1 - 250 of 114448 Next > Last > 250 Find place... Toggle Control Visibility

Results Models

Map Satellite

Indian Ocean Territory

Find place...

Select Layer...

115405 - layer values detail view Accepted Species - *Myrmicine_genus16 MG04*

115406 - layer values detail view Accepted Species - *Myrmicine_genus16 MG04*

115407 - layer values detail view Accepted Species - *Myrmicine_genus16 MG04*

B 115408 - layer values detail view Accepted Species - *Anochetus grandidieri*

C 115409 - layer values detail view Accepted Species - *Monomorium termitobium*

D 115410 - layer values detail view Accepted Species - *Plagiolepis alluaudi*

E 115412 - layer values detail view Accepted Species - *Monomorium versicolor*

D 115413 - layer values detail view Accepted Species - *Plagiolepis alluaudi*

D 115414 - layer values detail view Accepted Species - *Plagiolepis alluaudi*

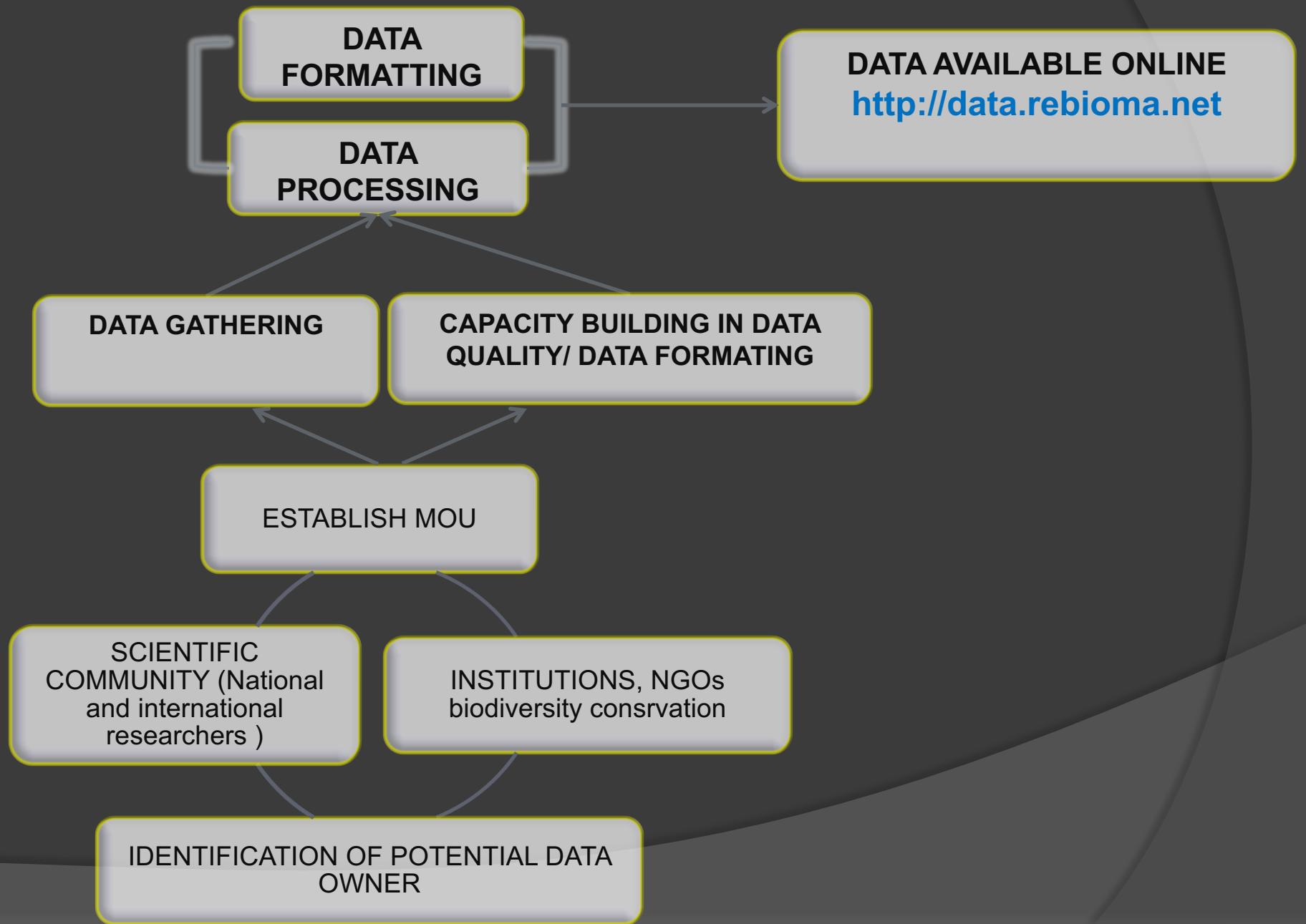
F 115415 - layer values detail view Accepted Species - *Tetramorium delagoense*

F 115416 - layer values detail view Accepted Species - *Tetramorium delagoense*

Map data ©2018 AfriGIS (Pty) Ltd, Google, ORION-ME 200 km Terms of Use

-06.69952513, 053.48144531

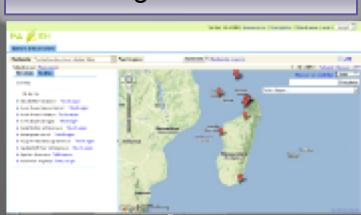
Strategy adopted



Structure et fonctionnalité du portail

Utilisateurs

Visualisation, Requêtes, Téléchargement



Planification
de la
conservation

Sciences

Professionnels



Applications

Modèles de distribution

Modélisation de la
distribution des espèces

Données évaluées

Oui
Non

Validation par les experts

Taxonomy
Review
Board



Taxonomie
Distribution

Qualité de données

Données validées

Notification

Oui
Non

Validation automatique
des données

Taxonomie, coordonnées
géographiques, date

Données invalides

Données

Données



Publications

Recherche



Specimens

Modèles de Distribution des espèces

- Modélisation automatique à partir de données publiques et privées certifiées par les experts
- Par le logiciel de modélisation MAXENT (<http://www.cs.princeton.edu/~schapire/maxent/>) est utilisé pour produire continuellement ces modèles de distribution

inputs: données d'occurrence évaluées fiables + données environnementales pour plusieurs époques.



Boophis madagascariensis



THOR HAKONSEN - 2010

Version 1.0.r1580 | tsikyrbetrano@yahoo.fr | Sign out | REBIOMA | Code site | Help | Acknowledgment | English ▾



occurrences Species explorer Data dashboard User Profiles

Search All Occurrences ▾ Public and private for taxon *boophis madagascariensis* Search Advanced Search | Shape file

Upload List Activities

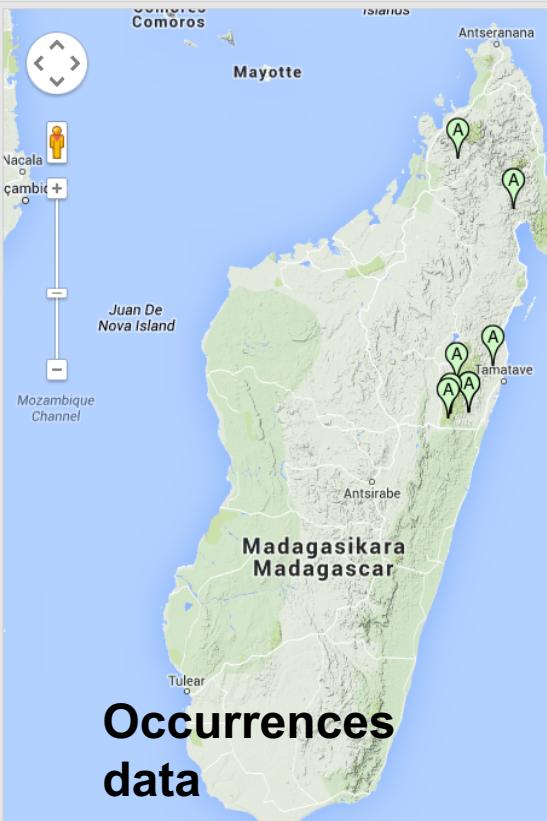
1 - 8 of 8

10 CSV

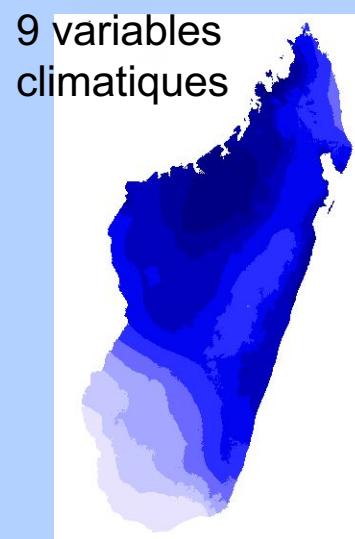
Select: All None

Results Models

- A 115332 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 222327 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 233457 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 261278 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 261324 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 261368 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 261406 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*
- A 261497 - [layer values](#) [detail view](#) Accepted Species - *Boophis madagascariensis*



+



+



Search All Reliable Reviews ▾ Public and private ▾ for taxon *boophis madagascariensis* Search Advanced Search | Shape file  1 - 8 of 8 10 

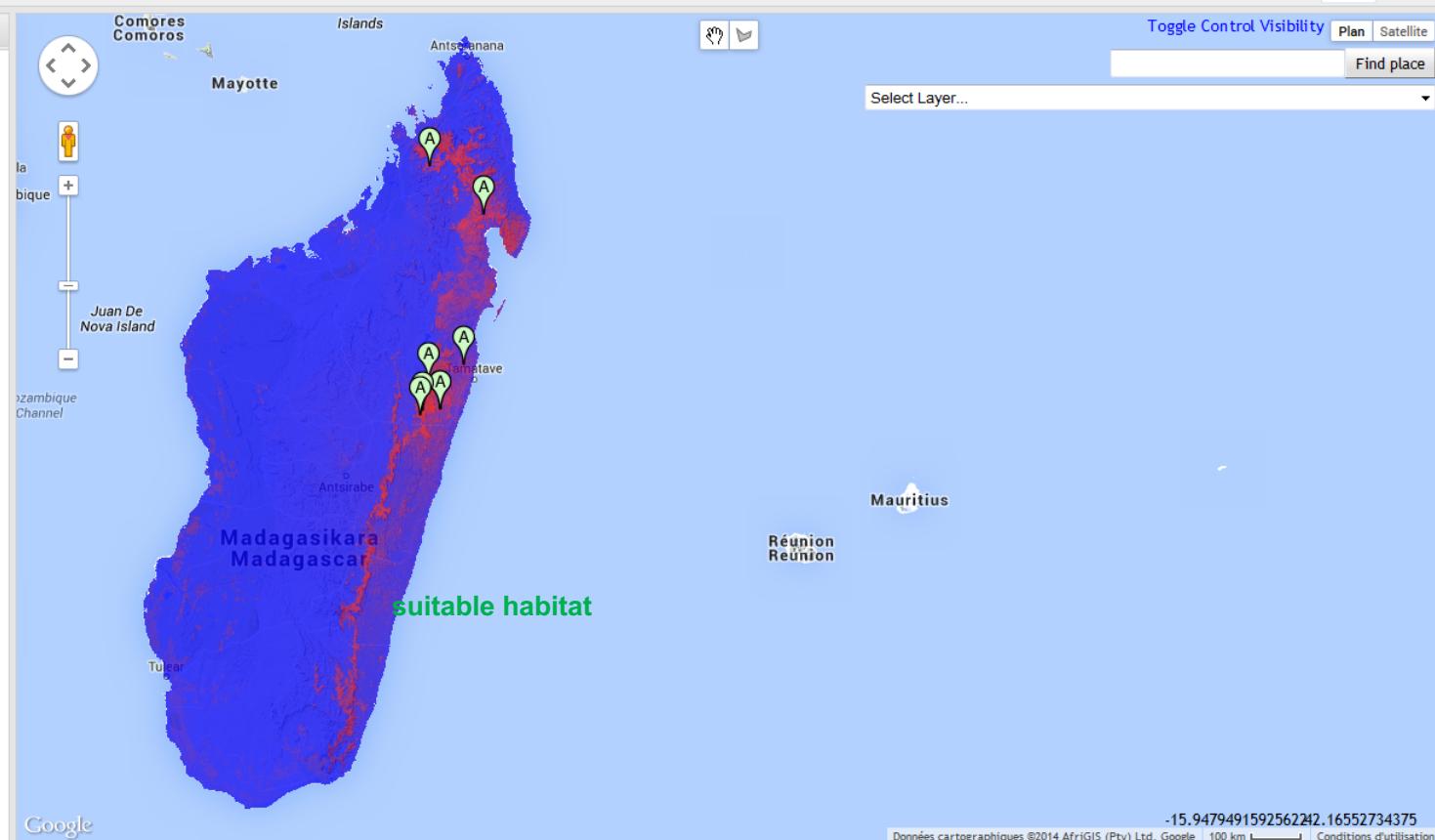
Select: All None

Results Models

Models

1 - 1 of 1

Boophis madagascariensis Download

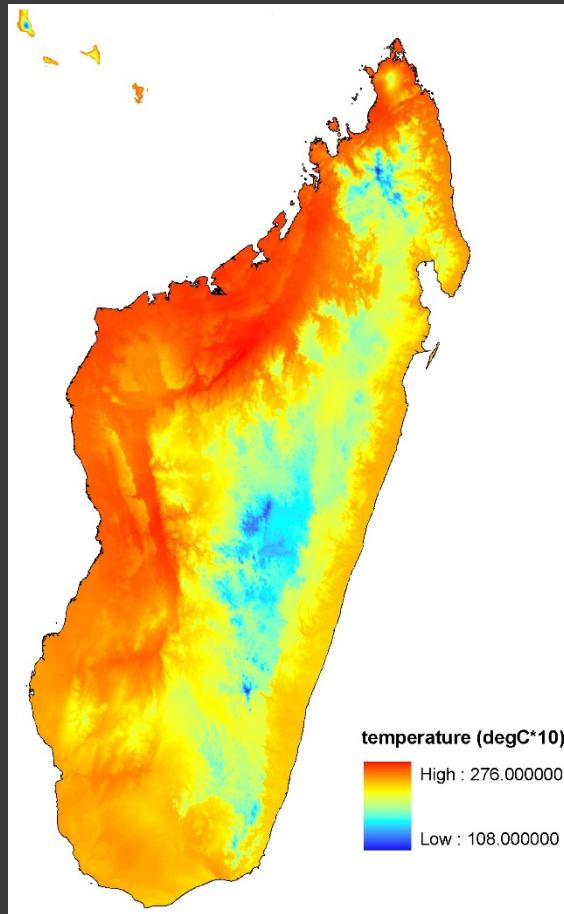
 2100_a2a Download 2100_b2a Download current Download

-15.9479491592562242.16552734375

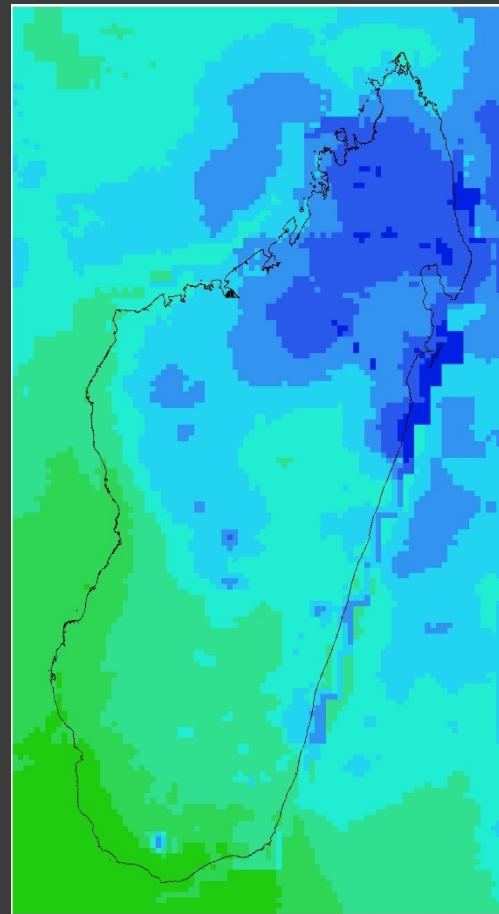
Données cartographiques ©2014 AfrigIS (Pty) Ltd, Google | 100 km  Conditions d'utilisation 

ENVIRONMENTAL DATA

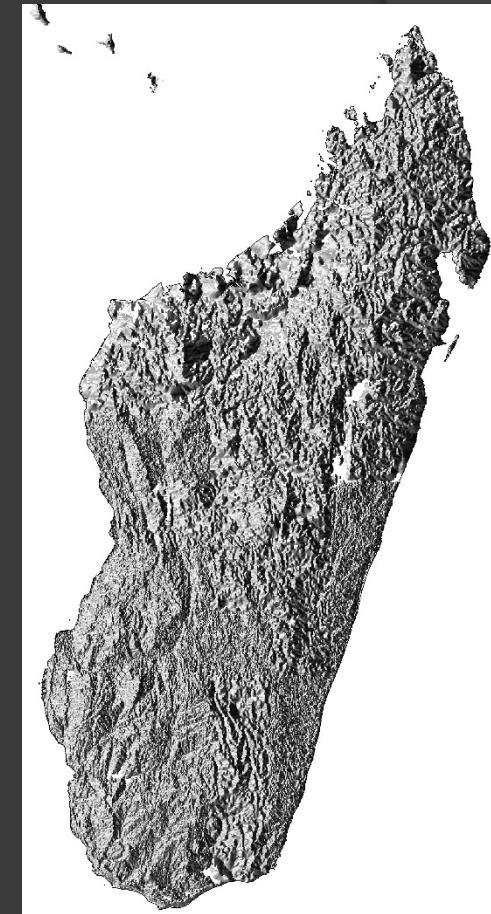
Common types



Mean annual temperature
Source: www.worldclim.org

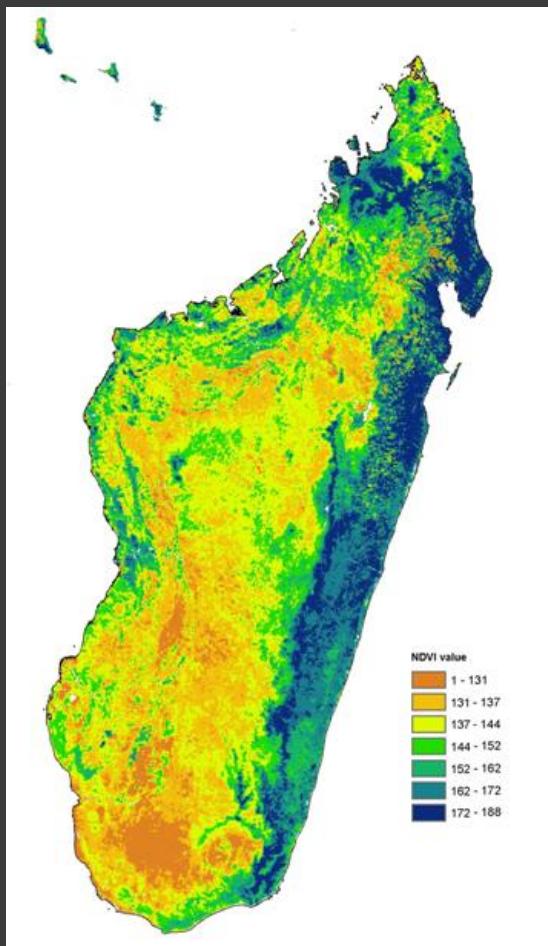


Mean annual precipitation
Source: NOAA FEWS

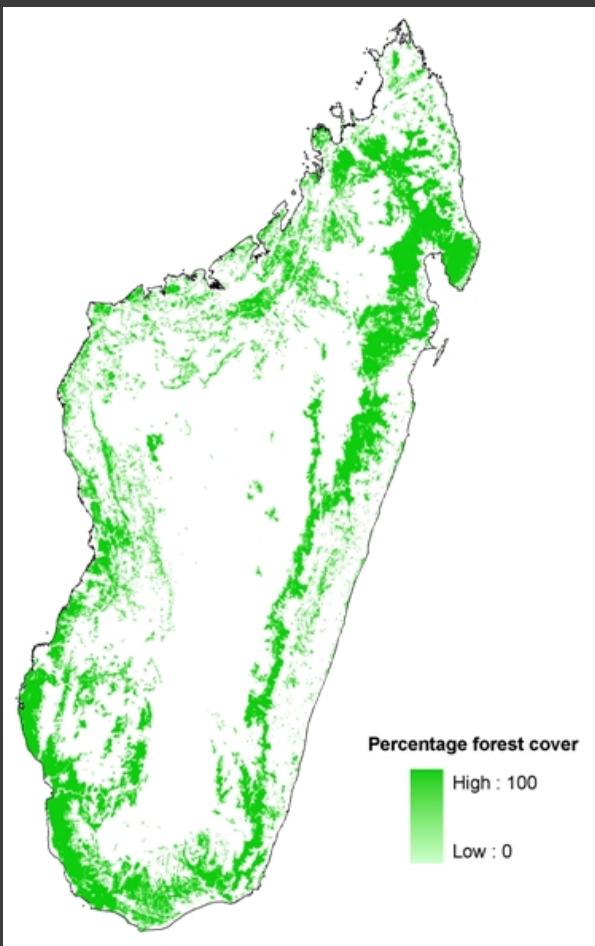


Aspect: East-West
Source: USGS Hydro1k

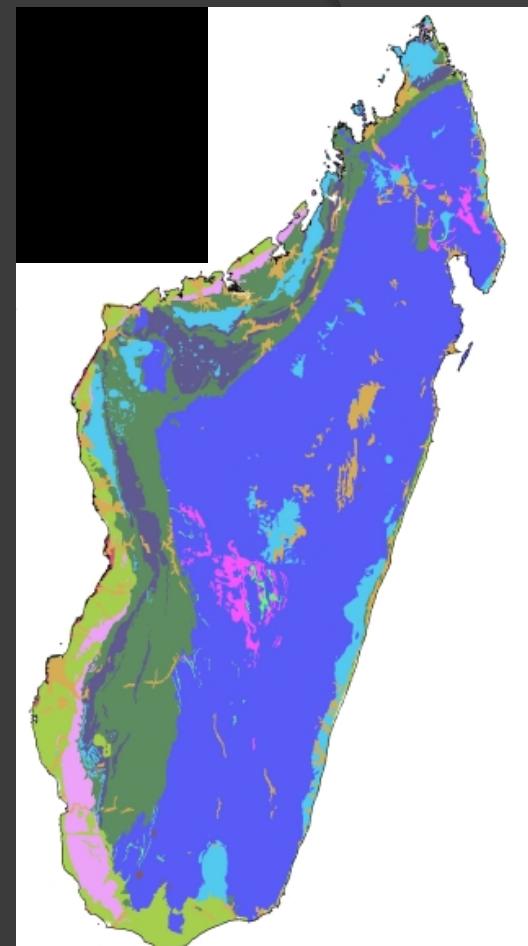
Common types



AVHRR NDVI
Source: NASA

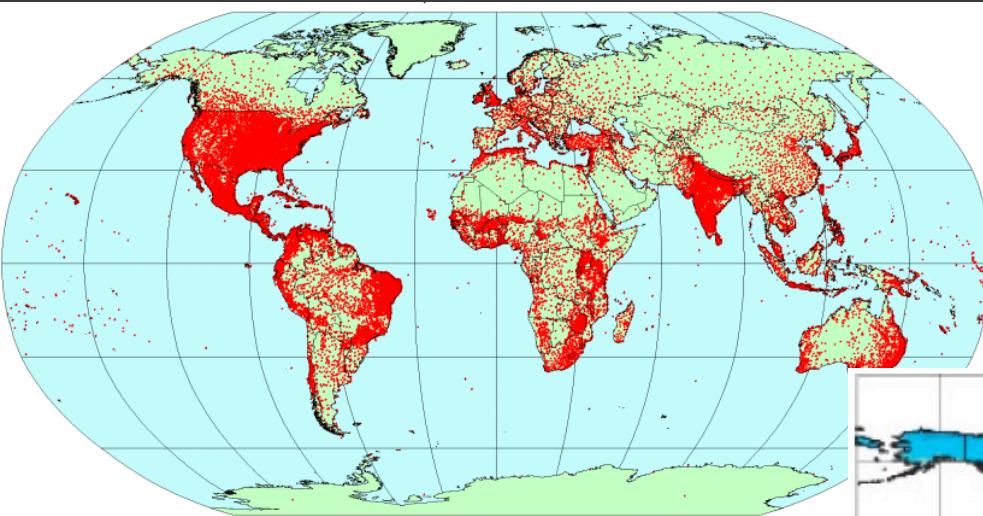


% forest cover
Source: IEFN and CI

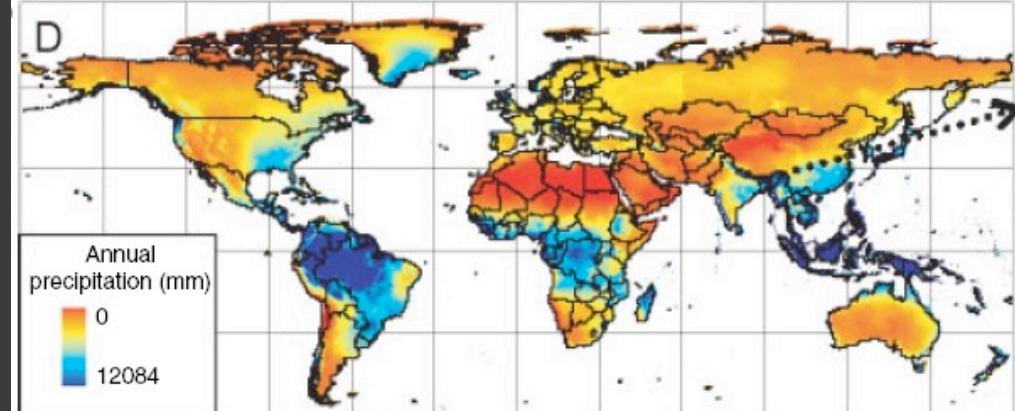
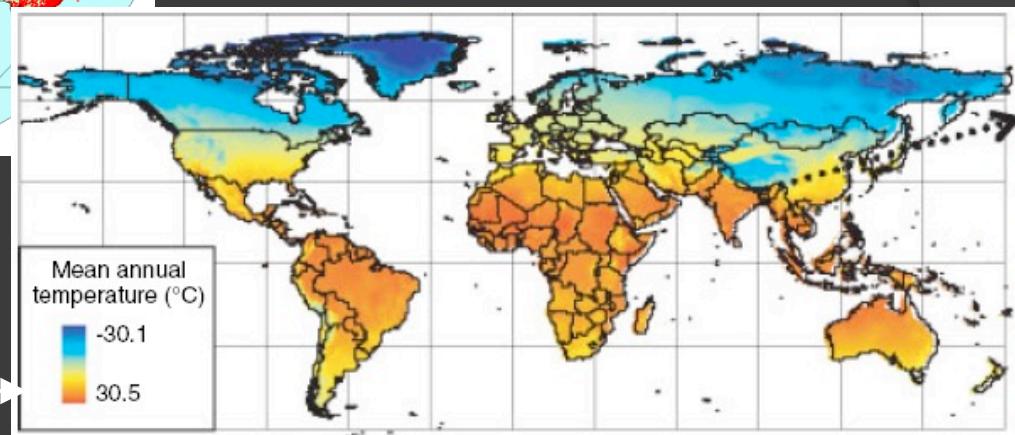
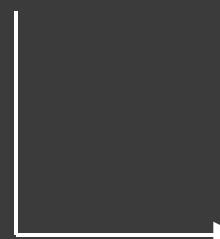


Geology
Source: Kew gardens
(note: categorical data)

Environmental Data: conversion to raster



Weather station
records
Vector point data;
(Robert Hijmans
et al. 2005)



Interpolated raster climate surfaces (WorldClim)

Some example sources of biological and environmental data for use in species' distribution modeling

Type of data	Source
<i>Species' distributions</i>	
- Data for a wide range of organisms in many regions of the world	Global Biodiversity Information Facility (GBIF): www.gbif.org
- Data for a range of organisms, mostly rare and/or endangered, and primarily in North America	NatureServe: www.NatureServe.org
<i>Climate</i>	
- Interpolated climate surfaces for the globe at 1km resolution (with bioclimatic variables)	WorldClim: http://www.worldclim.org/
- Scenarios of future climate change for the globe	Intergovernmental Panel on Climate Change (IPCC): http://ipcc-ddc.cru.uea.ac.uk/
- Reconstructed palaeoclimates	NOAA: http://www.ncdc.noaa.gov/paleo/paleo.html
<i>Topography</i>	
- Various atmospheric and land products from the MODIS instrument	NASA: http://modis.gsfc.nasa.gov/data/ http://edcimswww.cr.usgs.gov/pub/imswelcome/
<i>Soils</i>	
- Global soil types	UNEP: http://www.grid.unep.ch/data/data.php?category=lithosphere
<i>Marine</i>	
- Various datasets describing the world's oceans	NOAA: www.nodc.noaa.gov

Sites Web

- Portail de données: <http://data.rebioma.net>
- Atlas des APs: <http://atlas.rebioma.net>
- Institutionnelle: <http://www.rebioma.net>

Contact

- dimby@rebioma.net
- tsiky@rebioma.net
- rebioma@rebioma.net