

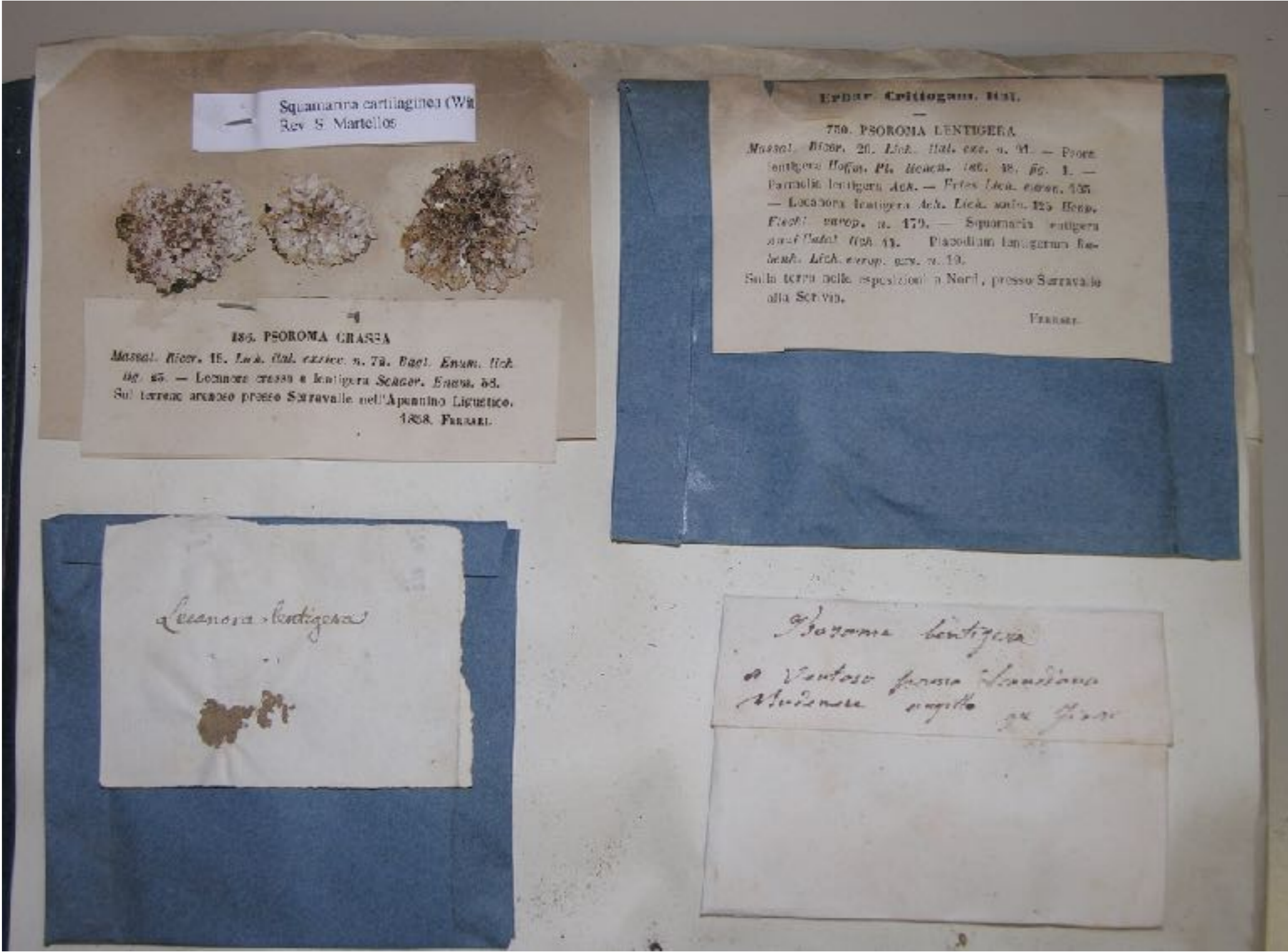
**Biological data:  
Primary biodiversity data**

**What primary biodiversity data are?**

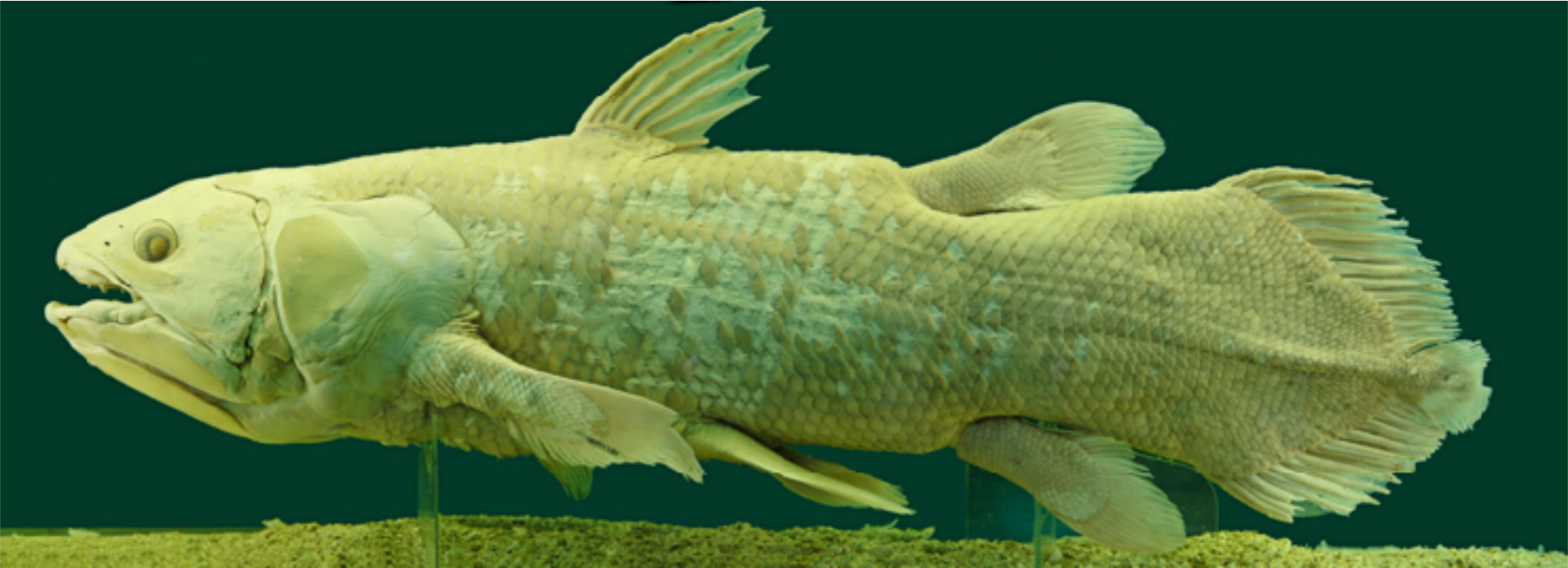
**Where do they come from?**

**Are they complex?**

Primary biodiversity data



# Primary biodiversity data



# Primary biodiversity data

CARRO UYUCA - DEPARTAMENTO DE MORAZAN - HONDURAS  
 5000 FT. Pine oak clearing - 12/15/46 - with  
 Dr. Paul C. Standley & Lewis O. Williams

3897 - *Distylium hondurense* Standl. & Wats. ex  
 Standl. - TREE - 15 m. det. Standley - (2)

3898 - *Salsola* - WOODY HERB - 1 1/2 m. Fls. Purple - (3)

3899 - MELASTOMACEAE - WOODY HERB - 1 1/2 m. Fls. pink

3900 - *Panicum cavirostris* - GRASS - WHITE - (1)

3901 - *Castilleja* - HERB - 1 m. (a)

3902 - *Ruellia* - Fls. yellow - (1)

3903 - *Valeriana sorbifolia* - (2)

3904 - COMPOSITE - WOODY HERB - 1 1/2 m. Fls. yellow - (2)

3905 - TREE - 20 m. 6000 FT. Cloud Forest - (2)

3906 - COMPOSITE - WOODY HERB - 2 m. Fls. white - (1)

3907 - *Salvia* - Fls. blue - common - (3)

3908 - *Salvia* - Fls. rose purple - (4)

3909 - *Salvia* - Fls. red - (1)

3910 - *Cestrum aurantiacum* - SARCO - 1 1/2 m.  
 Fls. orange - Fruits white - common - (2)

3911 - *Salvia* - Fls. pale lavender - (2)

3912 - *Eupatorium* - WOODY HERB - 2 m. Fls.  
 pale pink - (4)

RIO TABAZACA - VERAGUAS PROVINCE - PANAMA - 100 m.  
 Plants flowered in Tabaca 12/20/46

3913 - *Cycnocheilus ventricosus* - Mate flowers -  
 (SEE picture) ON PLANT ABOUT 8" HIGH - SWAY  
 TO FLEETLY - Small size of flower shows  
 to be correlated with size of plant - (1)

3914 - *Cycnocheilus ventricosus* - Mate flowers -  
 Small white flower in size of 3912 - (1)

VICINITY OF CAMPANA - 25 m. 6/28/46 -

3571 - *Miconia* - WOODY HERB - 2000 m. Fls. pink - Very showy - (18) (16)

3572 - *Oreocallis* - WOODY HERB - 2000 m. Fls. yellow - (6)

VICINITY OF LA CAMPANA - 25 m. 6/28/46 -

3573 - *Ruellia* - 1 m. - 8 m. Fls. white - very  
 fragrant - (30)

El Valle - North hills, 1000 m. 6/29/46

3574 - *Miconia Allanii* Standl. TREES - 70-80 m.  
 with buttressed trunks - flowers cream  
 white - (SEE picture of flowers & fruit.) (50)

3575 - *Oreocallis* - epiphytic - Plants resembling *Platanus*  
 flowers globose on a short stalk at top  
 base of the stem, pendant, of 5-6  
 dull purple flowers reminiscent of  
*Cyrtanthera* except that the  
 sepals are not connate at their tips -  
 lip trilobate - possibly in *veraguensis*  
 group. (SEE picture.) (2)

3576 - *Aechmea Allanii* L.B. Smith - Fruit capsule  
 black - leaves grey white - common - (1)

El Valle - Floor - 600 m. 6/29/46

3577 - *Chimaphila* - WOODY HERB - (1)

NOTES copied 9/11/46  
 to Herbar. Re-Sent 3/5/68

Primary biodiversity data should answer 4 questions:

1. What was collected / observed (the name of the organism)?
2. Where was it collected / observed (locality, geo-referenced or not)?
3. When was it collected / observed?
4. By whom was it collected / observed?

While normally observations made in the field do contain these 4 information alone, even if sometimes completed by some notes on the site, natural history collections specimens often contain a wealth of other information. While modelling the climatic niche of an organism, we need coordinates of occurrence points. However, when we are trying to define the systematic position of a taxon, many other information could be relevant.

Biological collections are the source of PBD, and include:

- Collections of living organisms
- Natural history collections
- Botanical or zoological observations

Especially in the second case, they are an important source of **falsifiable information**.

A statement is **falsifiable** if some observation can prove it to be false.

The concept was introduced by Karl Popper in his book *The Logic of Scientific Discovery*, in 1934, as the cornerstone of his view of science as critical rationalism.

Specimens stored in natural history collections are falsifiable, in the sense that any hypothesis made on them can be proven false. This is doable by viewing, or physically accessing specimens.

Thus, the relevance of storing specimens when doing any research, from molecular phylogeny (in which specimens are stored in NHM collections, and sequences are stored in online data banks) to taxonomy (in which specimens, especially type specimens, which are the voucher of a new taxon name, are stored in NHM collections).

In the case of primary biodiversity data, accessing a specimen could allow especially the falsification of the name (i.e., one can verify whether the identification made by previous researchers is correct). However, other data can potentially be falsified as well, even if the process is often indirect.

Is an observation falsifiable?

Yes. No. Maybe.

It can be potentially falsified indirectly, and directly.

Indirectly, one can check whether the observation falls into the known range of a species. If yes, it could be classified as correct. However.... If one reports a seagull, inside the known range of the species, while observing a blackbird, the observation is incorrect, even if, by indirect falsification, it could be classified as correct.

Directly, one can go in the field and check whether the observation is correct, i.e. whether the reported organism is present in that very location.

However, if the observation is old (months, years...), depending on the lifespan, and phenology of an organism, falsification could not be possible. Plus, if the observed organism is an animal, it could have moved, thus not be present in that very location, even if the observation was actually correct.

Furthermore, even if one goes in the field, and observes the reported organism in the same location, can it be stated that the observation was correct?



Let's make a simple example.

I go in the field, and report a seagull.

I am not an ornithologist, thus I mistake a blackbird for a seagull.

After a little time, one is willing to model the distribution of seagulls in Trieste, and wants to check my observation.

In the field, in the location I reported, he observes a seagull. Thus, he states that my observation was correct.

But it was not.....

Plus, even if me and the one who verifies the observation observe the same organism, was it the same individual?



**Big problem:**

**To be used, data should exist in the  
digital domain....**

## Primary biodiversity data: The GBIF

---

---

General	GBIF	<a href="http://www.gbif.org">www.gbif.org</a>
General	Map of Life	<a href="https://mol.org/">https://mol.org/</a>
General	LifeMapper	<a href="http://lifemapper.org/">lifemapper.org/</a>
General	IUCN Red List	<a href="http://www.iucnredlist.org/">www.iucnredlist.org/</a>
Herps	HerpNET	<a href="http://herpnet.org/">herpnet.org/</a>
Mammals	MaNIS	<a href="http://vertnet.org/">vertnet.org/</a>
Marine species	OBIS	<a href="http://www.iobis.org/">www.iobis.org/</a>
Amphibians	AmphibiaWeb	<a href="http://amphibiaweb.org/">http://amphibiaweb.org/</a>
Birds	ORNIS	<a href="http://ornisnet.org">http://ornisnet.org</a>
Birds	Bird Life	<a href="http://www.birdlife.org/">www.birdlife.org/</a>
Plants	Atlas Flora Europaea	<a href="http://www.luomus.fi/en/database-atlas-florae-europaeae/">www.luomus.fi/en/ database-atlas-florae-europaeae/</a>
Plants	BIEN	<a href="http://bien.nceas.ucsb.edu/bien/">http://bien.nceas.ucsb.edu/bien/</a>
Central America	REMIB	<a href="http://www.conabio.gob.mx/remib_ingles/doctos/remibnodosdb.html?">www.conabio.gob.mx/remib_ingles/ doctos/remibnodosdb.html?</a>
Brazil	SpeciesLink	<a href="http://splink.cria.org.br/">http://splink.cria.org.br/</a>

---

---

# Primary biodiversity data: The GBIF

GBIF | Global Biodiversity Information Facility

## Free and open access to biodiversity data

Get data | How-to | Tools | Community | About

Login

OCCURRENCES | SPECIES | DATASETS | PUBLISHERS | RESOURCES

Search

WHAT IS GBIF? | ABOUT GBIF ITALY

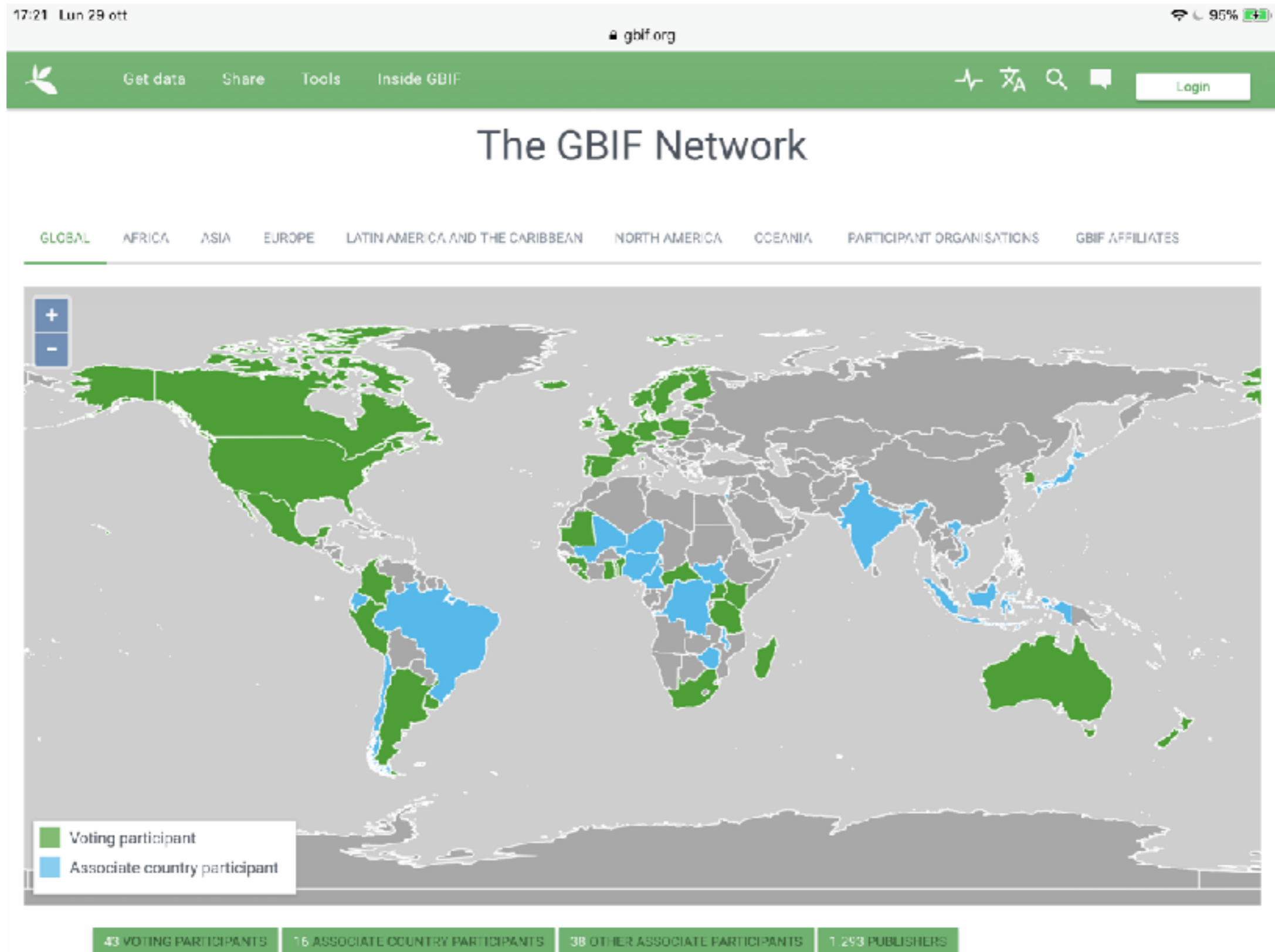
Mustard surgeonfish (*Acanthurus gurtetus*) observed near Namukulu, Niue by emr\_auckland. Photo via iNaturalist (CC BY-NC 4.0)

Occurrence records	Datasets	Publishing institutions	Peer-reviewed papers using data
1.403.571.225	51.736	1.587	4.362

**News**

- Call for nominations to the 2020 GBIF Young Researchers Award
- Relying on biodiversity science to inform art history
- 2020 Ebbe Nielsen Challenge seeks open-data innovations for biodiversity
- Biodiversity Summit 2020 postponed

# Primary biodiversity data: The GBIF



## Primary biodiversity data: The GBIF

The GBIF - some figures [at 2021/03/16, 5:52 pm]:

1.662.149.688 occurrence records

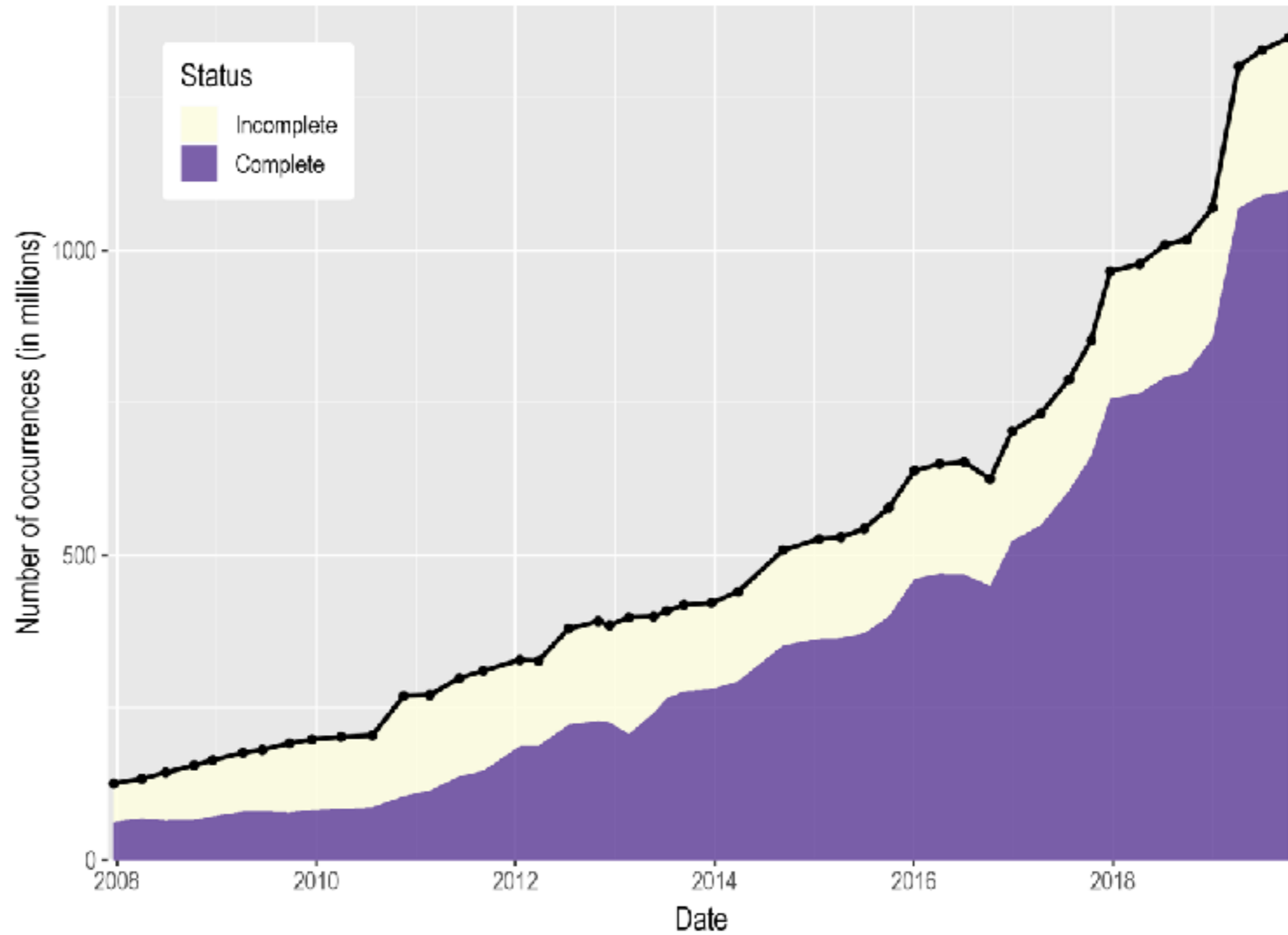
57.008 datasets

1.655 publishing institutions

5.579 peer reviewed papers using GBIF data

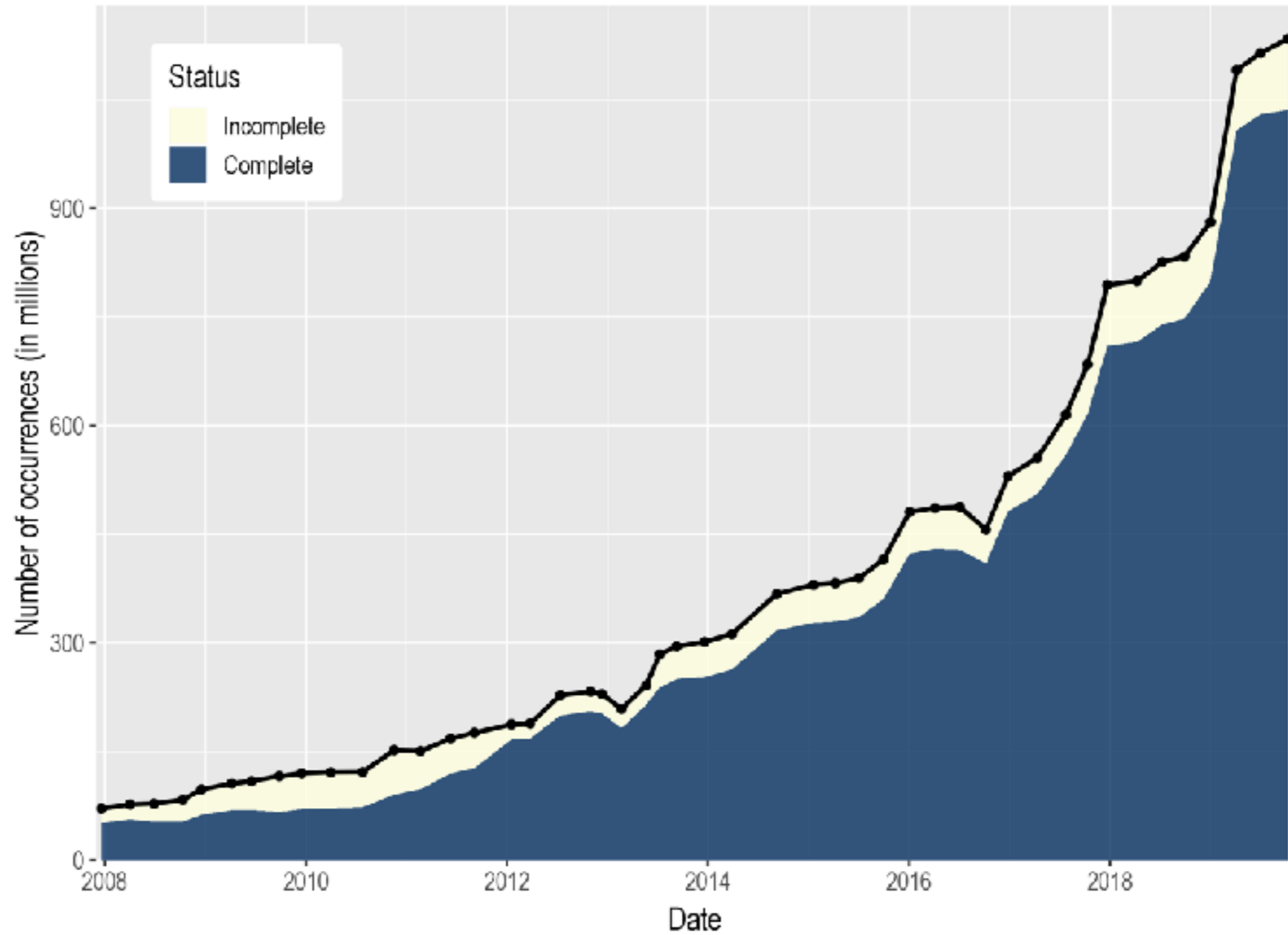
# Primary biodiversity data: The GBIF

Complete species occurrence records accessible through GBIF over time



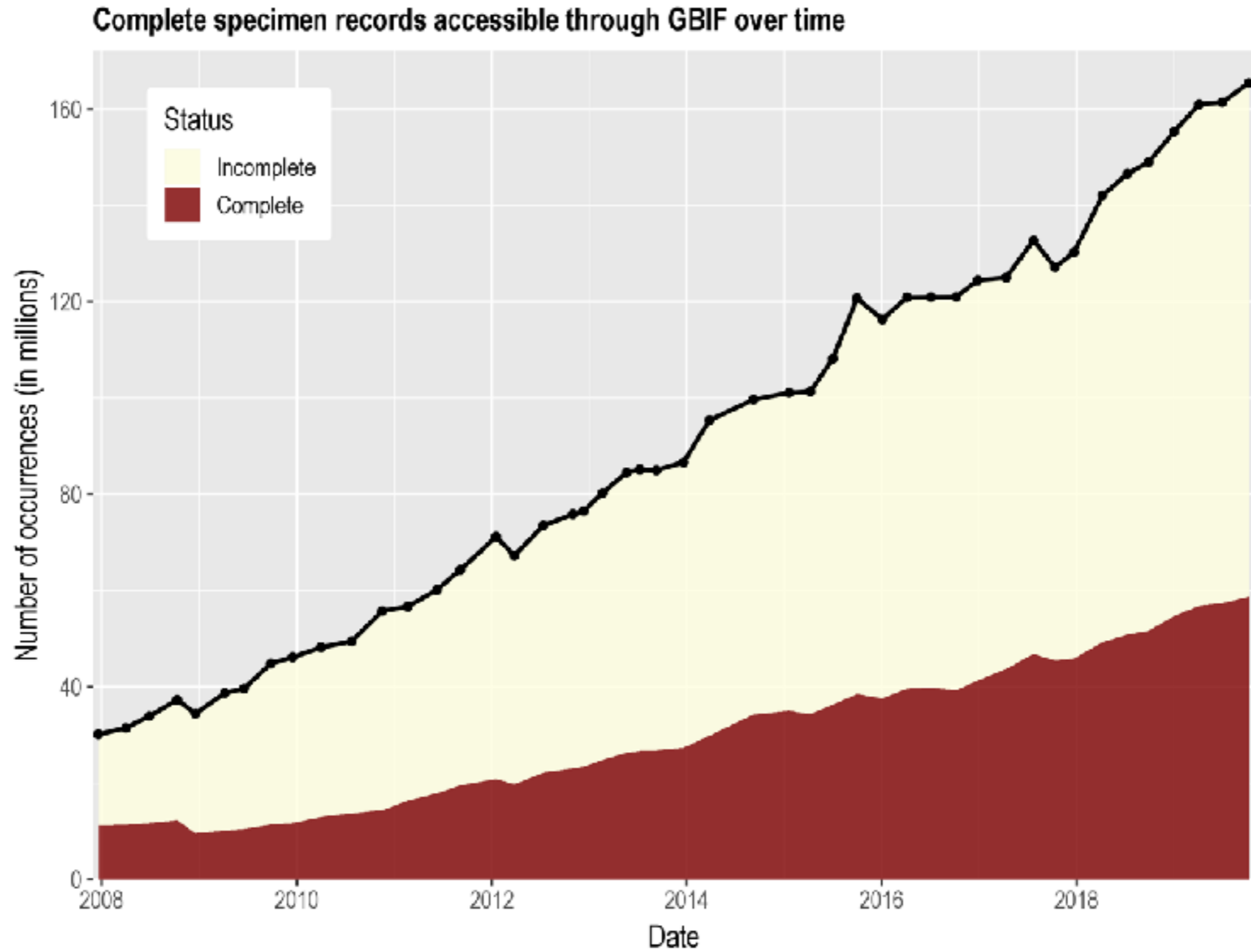
# Primary biodiversity data: The GBIF

Complete observation records accessible through GBIF over time





# Primary biodiversity data: The GBIF



## Primary biodiversity data: The GBIF

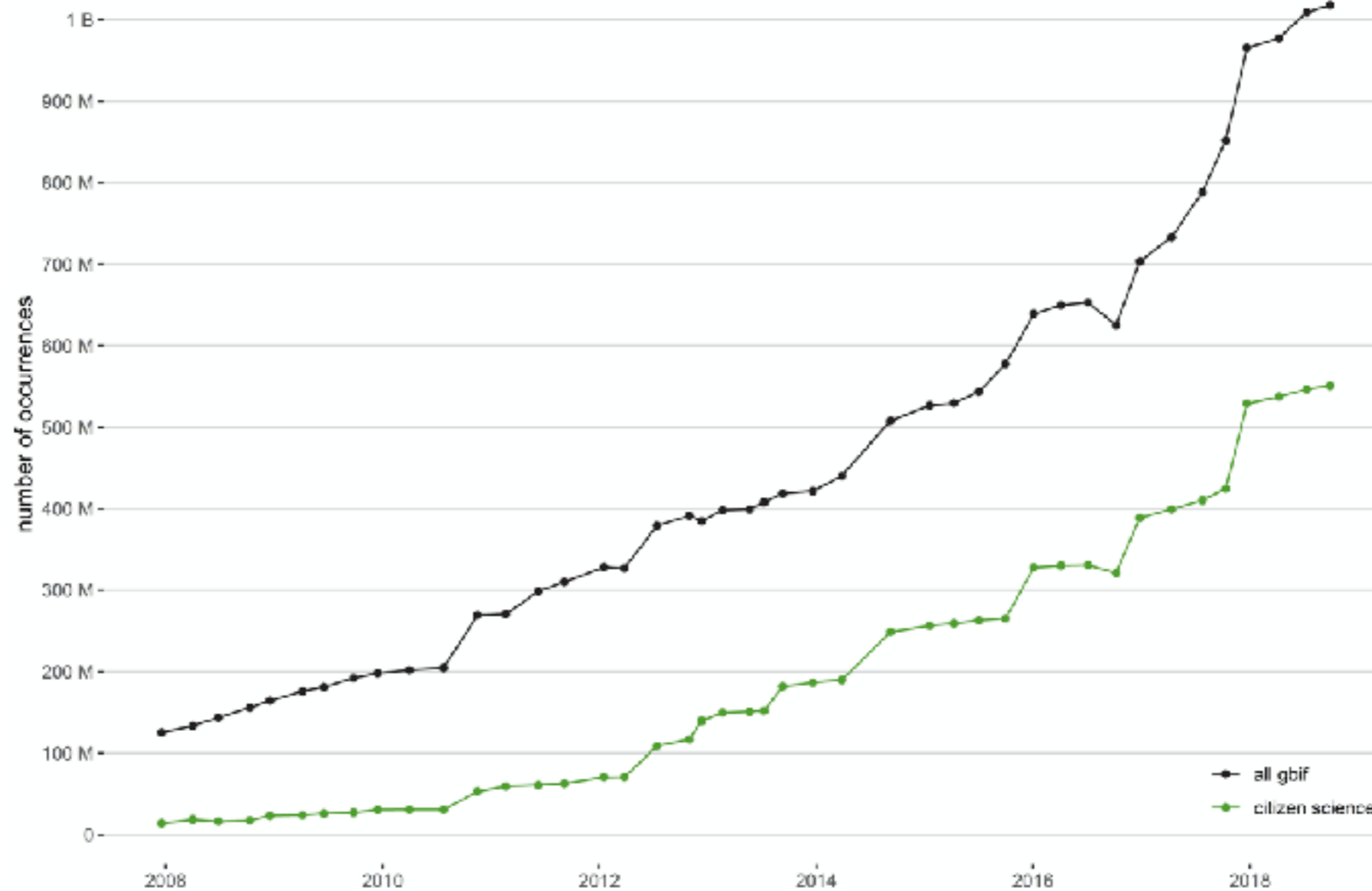
Observation records are the most complete, since they often are natively in the digital domain, especially when they come from citizen science activities.

Basis of record		^
<input type="checkbox"/>	Observation	19.653.975
<input type="checkbox"/>	Machine observation	11.912.825
<input type="checkbox"/>	Human observation	1.105.925.840
<input type="checkbox"/>	Material sample	25.326.297
<input type="checkbox"/>	Literature	219.987
<input type="checkbox"/>	Preserved specimen	166.087.620
<input type="checkbox"/>	Fossil specimen	9.029.568
<input type="checkbox"/>	Living specimen	1.637.949
<input type="checkbox"/>	Unknown	17.147.671

Furthermore, when they are not, their digitization is often far easier than for natural history collection specimens.

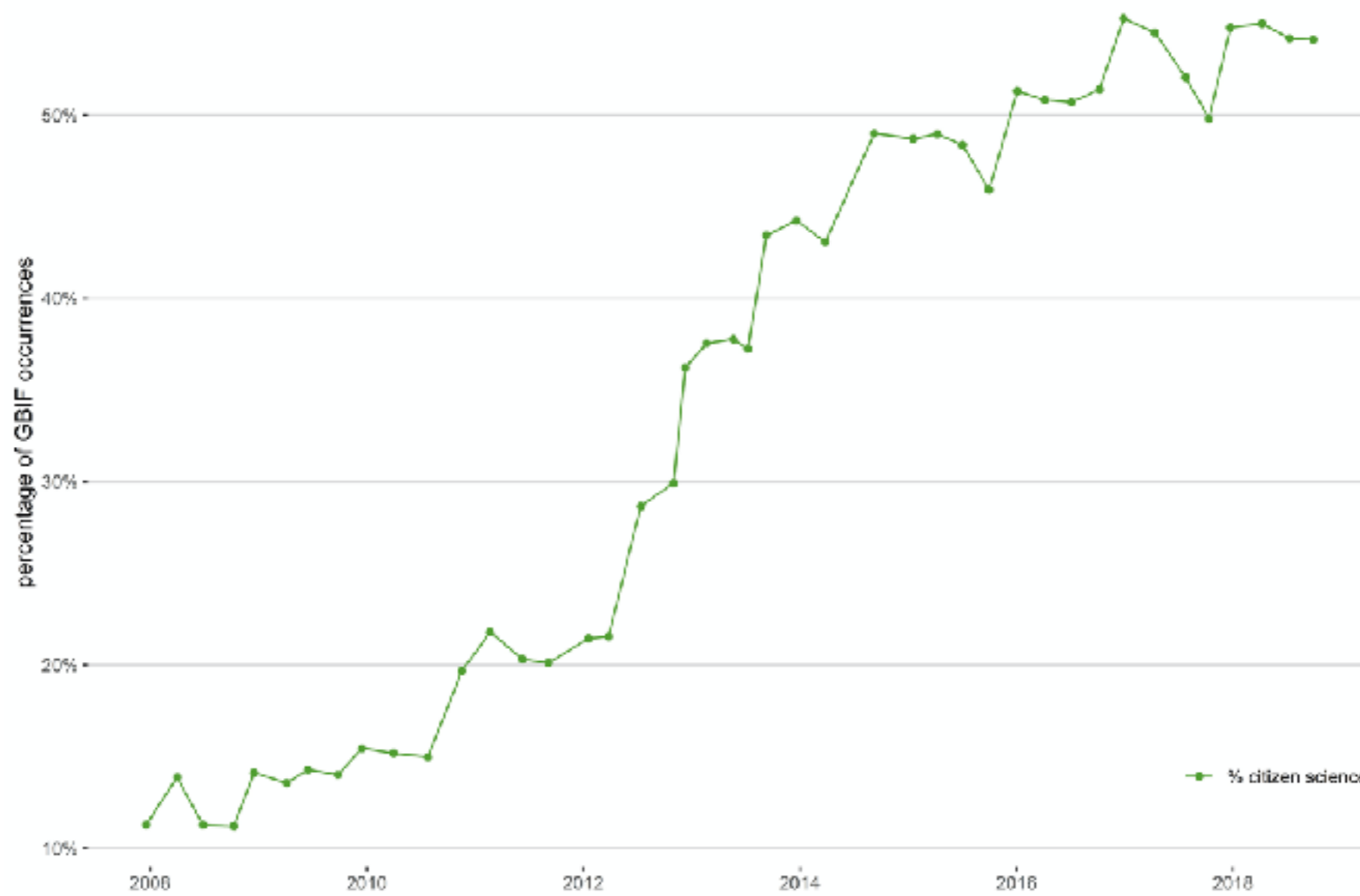
# Primary biodiversity data: The GBIF

Just to give you an idea of the relevance of *citizen science* as far as GBIF data are concerned.....



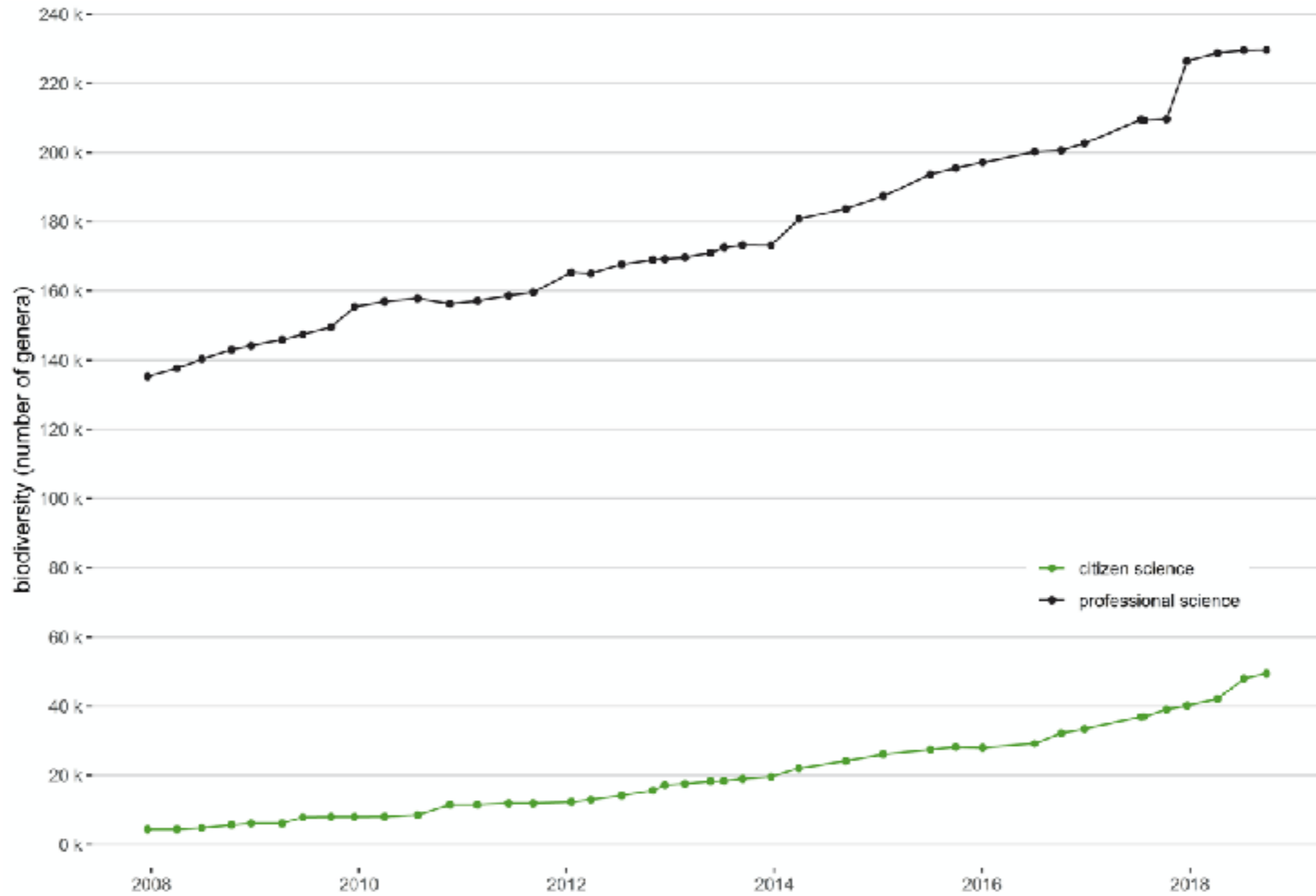
# Primary biodiversity data: The GBIF

50% of occurrence records on GBIF are citizen science observations



# Primary biodiversity data: The GBIF

However....



# Primary biodiversity data: The GBIF

There are > 8 000 000 seagull occurrence records



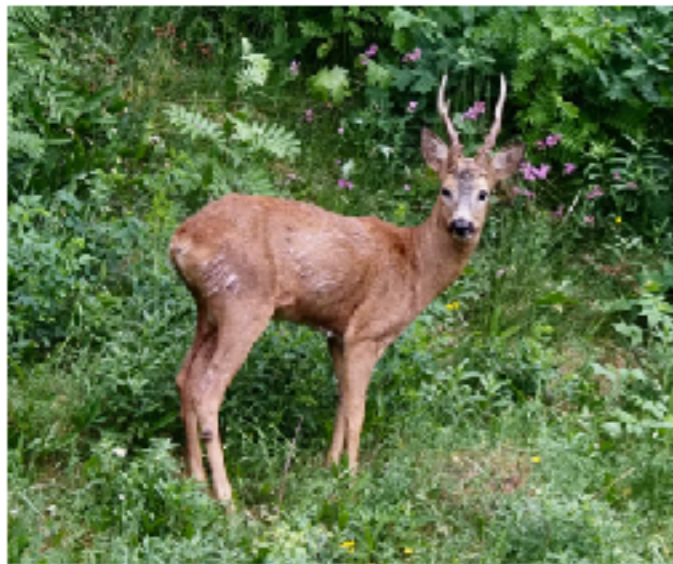
*Larus argentatus* subsp. *argenteus* by G.Droege via Botanic Garden and Botanical Museum Berlin-Dahlem. Photo licensed under [CC BY-SA 3.0](#)

Top insect is the red admiral butterfly 200k records



*Vanessa atalanta* by W.-H. Kusber via BoBO - Botanic Garden and Botanical Museum Berlin-Dahlem Observations. Photo licensed under [CC BY-SA 4.0](#)

Top mammal is the roe deer 140k records



*Capreolus capreolus* by Trine Brevig via the Norwegian Species Observation Service. Photo licensed under [CC BY 4.0](#)

Top plant is the common nettle 100k records



*Urtica dioica* subsp. *dioica* by Peter de Lange via iNaturalist. No copyright.

**GBiF data:  
should we trust them  
(human observations especially)?**

**Yes.**

**No.**

**Maybe...**



**Examples of biases  
due to low quality,  
or poor PBD**

In a paper of 2015, Medone et al. Described the possible switch in climatic niche for two insects of the family Triatominae, which are vectors for the parasite *Trypanosoma cruzi*, which is a relevant disease in South America.

The authors, starting from expert accessed distributional maps, derived a great amount of geo-referenced occurrence points for the two species, and developed maps of current and future (2050) distribution, with the aim of understanding which areas will be more affected by disease after the impact of foreseen climate changes.

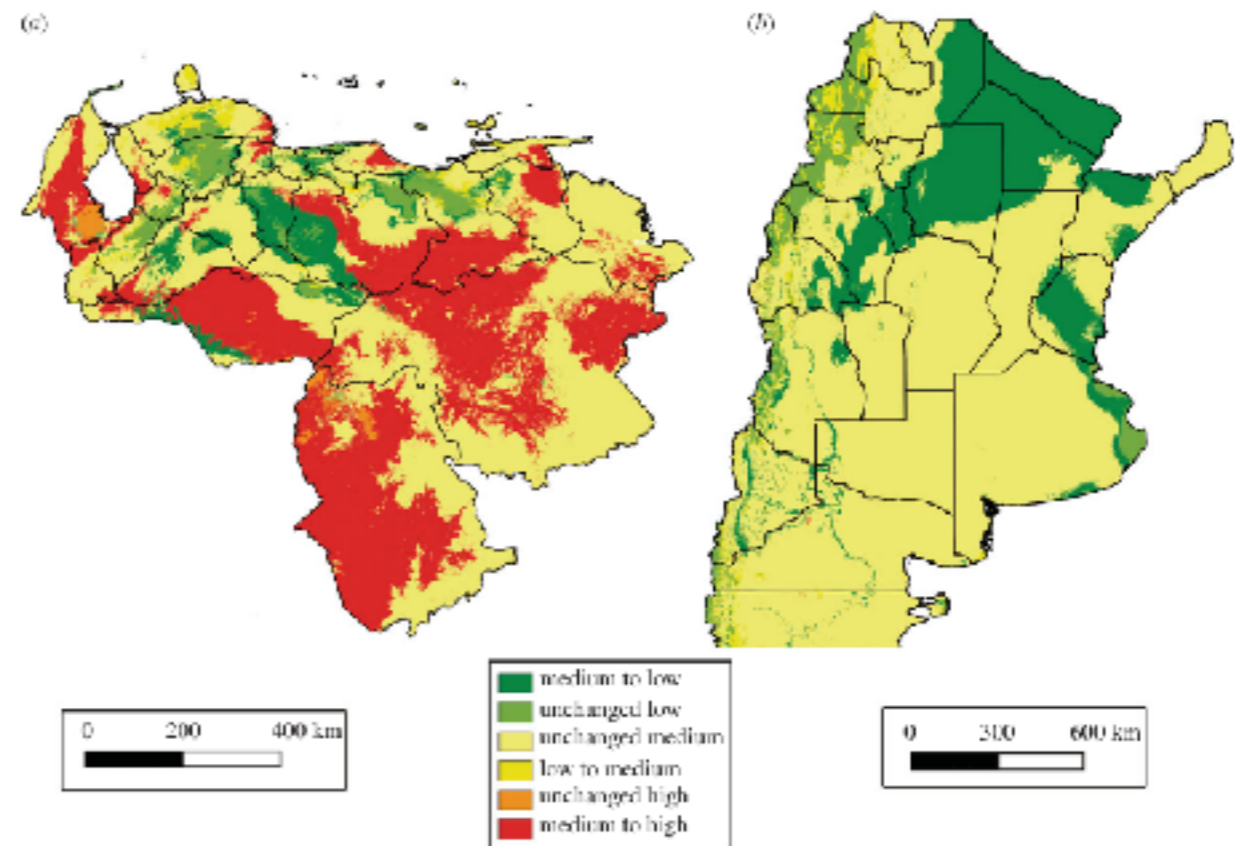


Figure 1. Climatic suitability changes from current to 2050 conditions for (a) *R. proflava* and (b) *T. infestans*. Colours indicate climatic suitability transitions between the three main suitability categories (low, medium and high). (Online version in colour.)

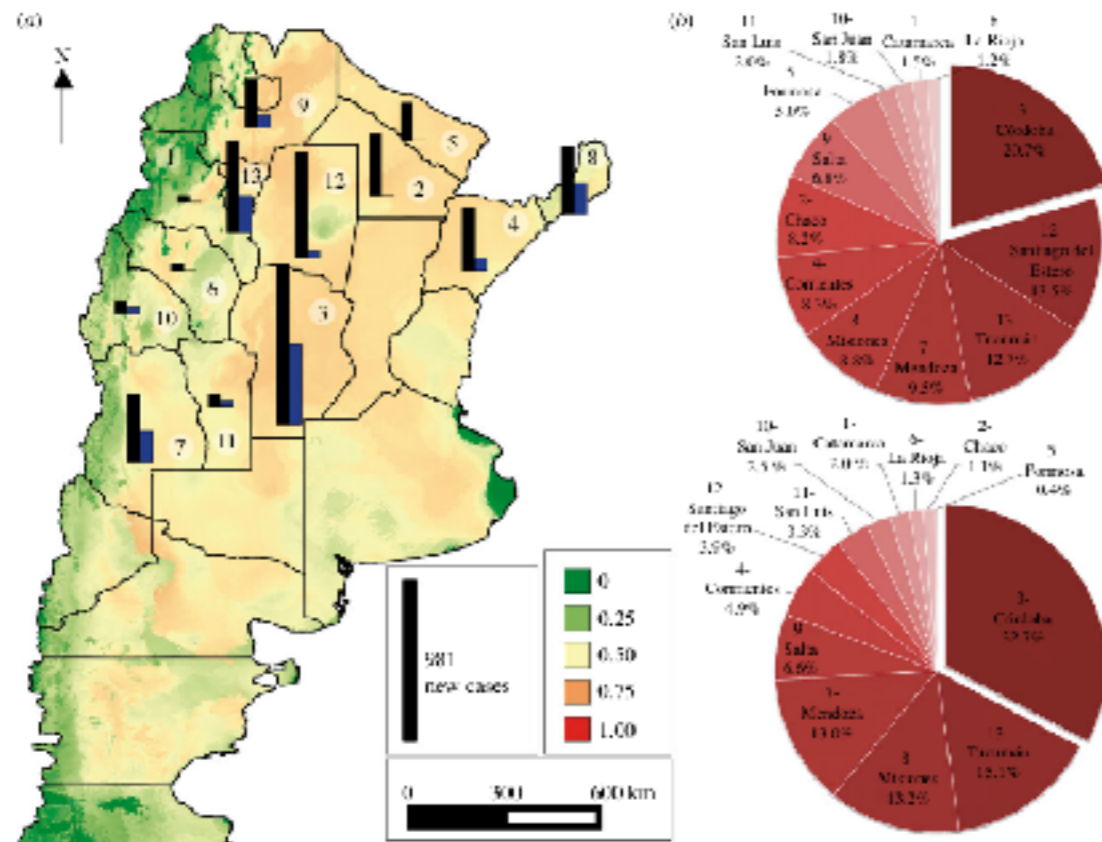


Figure 3. Current climatic suitability (base map green to red/grey to black colours) and predicted number of new human cases of *Y. enterocolitica* infection (current: black bar, 2050: blue/grey bar) in Argentina. Provinces were numerated (white circle on the map) to facilitate the identification on the pie charts (number preceding name). Pie charts indicate the 'relative contribution' (%) of each province to the total number of new predicted cases over all high transmission risk provinces. Upper pie chart: current conditions; lower pie chart: 2050 conditions. (Online version in colour.)

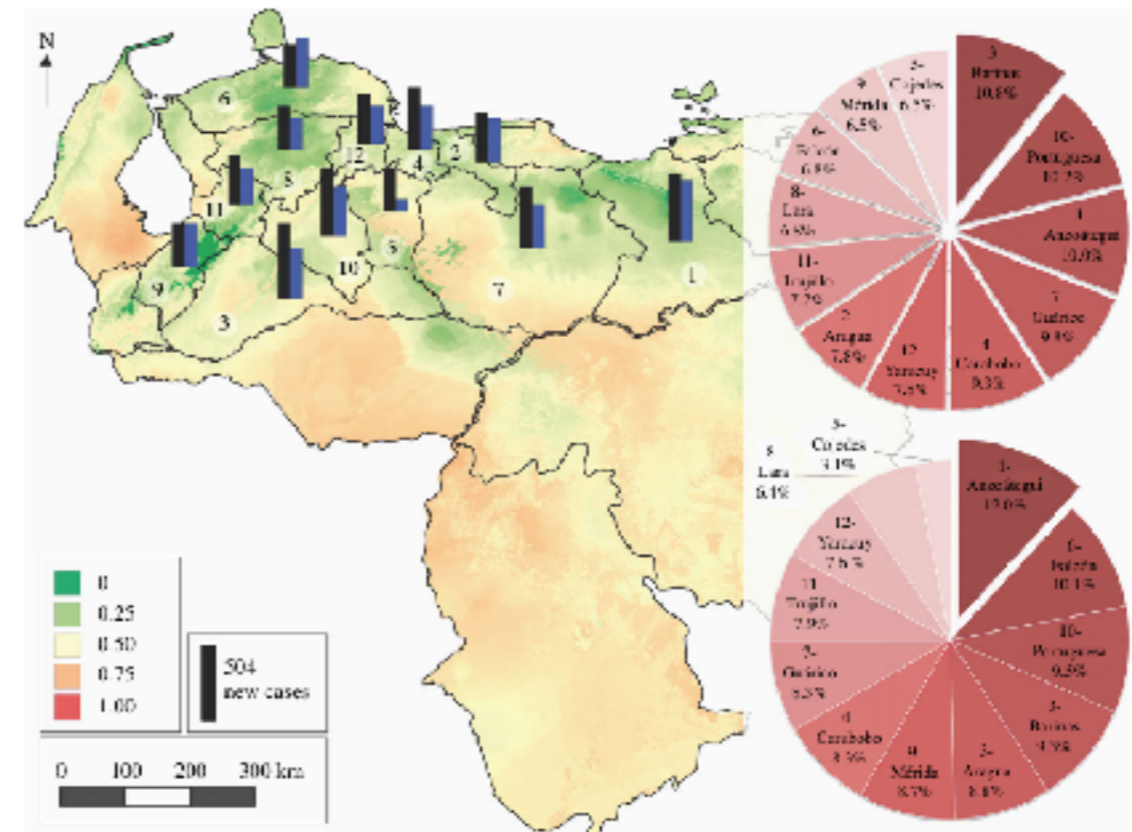


Figure 7. Current climatic suitability (base map green to red/grey to black colours) and predicted number of new human cases of *Y. enterocolitica* infection (current: black bar, 2050: blue/grey bar) in Venezuela. States were numerated (white circle on the map) to facilitate the identification on the pie charts (number preceding name). Pie charts indicate the 'relative contribution' (%) of each province to the total number of new predicted cases over all high transmission risk states. Upper pie chart: current conditions; lower pie chart: 2050 conditions. (Online version in colour.)

It is obvious that the presence points obtained by the authors are “pseudo-presences”.

Since they derive from maps depicting not the **area of occupancy** of the two species, but the **extent of occurrence**, they hardly can be used for making assumptions on the actual distribution of a taxon.

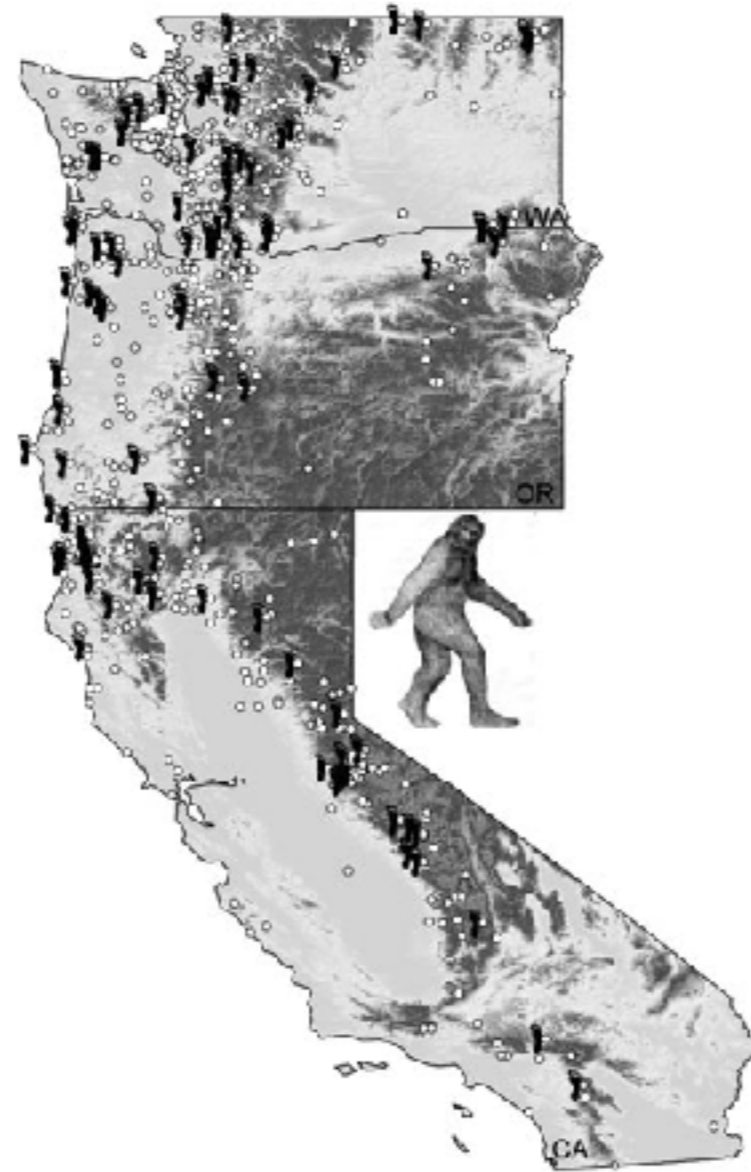
Furthermore, deriving considerations on the possible spread of parasitic infections on the basis of such data is far more than a long shot.

To highlight this issue, together with the importance of correct identification of each specimen (or observation), Lozier et al., in 2009, made an interesting exercise modelling the niche of a crypto-species, the Sasquatch, in western North America.

They collected all the data of sightings of the “animal” and of its footprints, geo-referenced them *a posteriori* with the highest degree of confidence as possible, and produced models for present and future distribution as an effect of climate change.

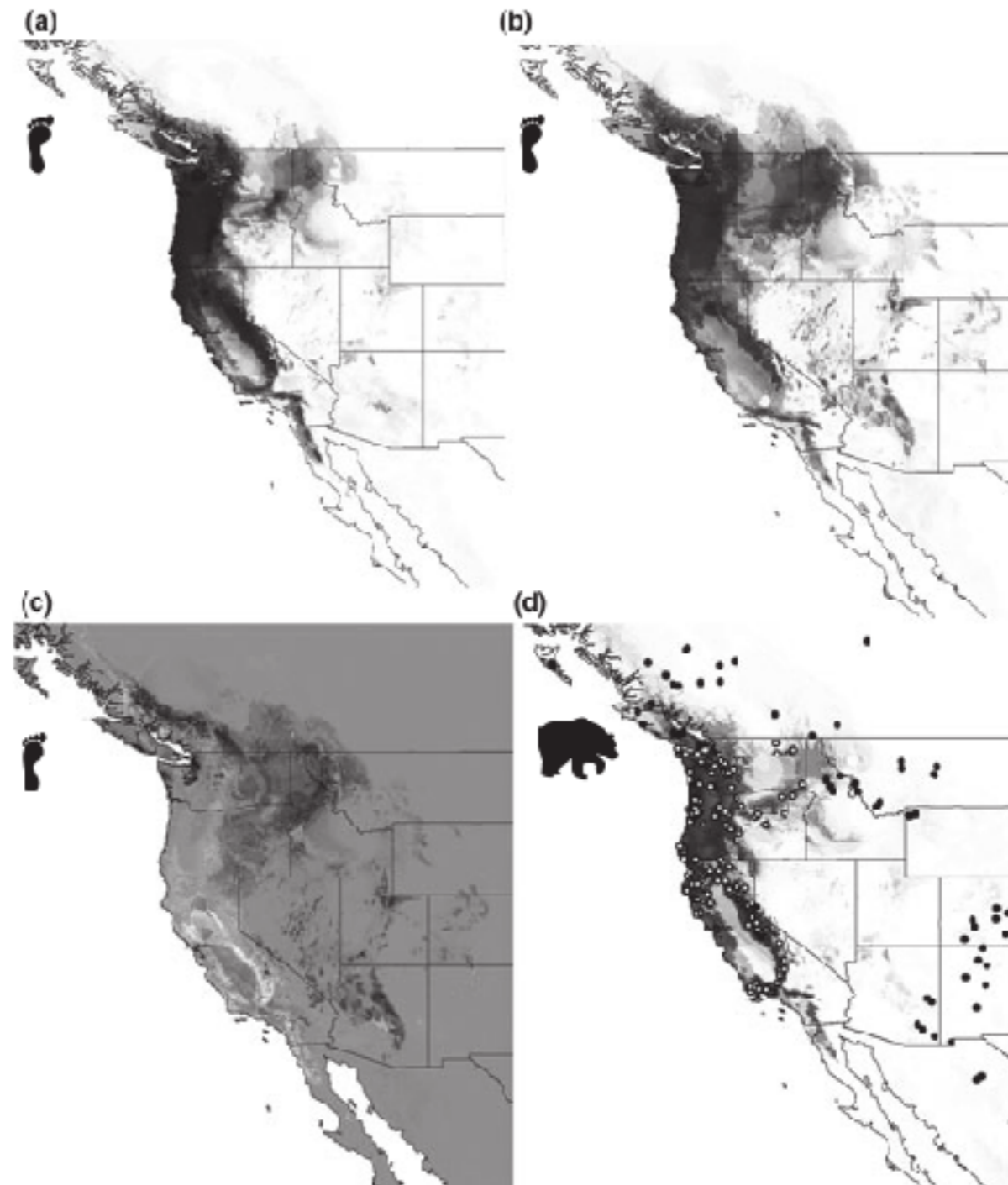
Furthermore, they compared the models with those of the black bear, animal which often is misidentified as Sasquatch.





**Figure 1** Map of Bigfoot encounters from Washington, Oregon and California used in the analyses. Points represent visual/auditory detection, and foot symbols represent coordinates where footprint data were available. Shading indicates topography, with lighter values representing lower elevations.

**Figure 2** Predicted distributions of Bigfoot constructed from all available encounter data using MAXENT (a) for the present climate and (b) under a possible climate-change scenario involving a doubling of atmospheric CO<sub>2</sub> levels. Results are presented for logistic probabilities of occurrence ranging continuously from low (white) to high (black). Differences between (a) and (b) are shown in (c), with whiter values reflecting a decline in logistic probability of occurrence under climate change, darker values reflecting a gain, and grey reflecting no change. A predicted distribution of *Ursus americanus* in western North America under a present-day climate is also shown (d). White points indicate sampling localities in California, Oregon and Washington taken from GBIF ( $n = 113$  for training, 28 for testing; compare with Fig. 1) used for the MAXENT model with shading as in (a) and (b); black points indicate additional known records not included in the model.



The authors were not willing to criticize the use of ENMs as a whole, but to evidence that data quality can badly affect the results and their interpretation.

In particular, they evidenced that:

- careful scrutiny of specimen records should be encouraged
- assessing specimens should be done whenever possible (**taxonomic impediment**)
- identifications should be made by expert taxonomists (**taxonomic impediment**), and specimens identified by taxonomists should be preferred when available (**specimens > observations**)
- all efforts should be made to ensure taxonomic accuracy when digitizing specimen data (**taxonomic impediment**)

As a conclusion, they state that today more than in the past well trained taxonomists are fundamental.



An interesting study by Romero et al., 2014, describe the effect of changing taxonomy on two newt species, *Triturus pygmaeus* and *T. marmoratus*, both known to occur in Spain. The two were considered a unique species, but are now treated as separate species thanks to DNA evidences.

It is interesting to notice how the models are affected by a separate treatment vs. the modelling of the two as a single taxon.

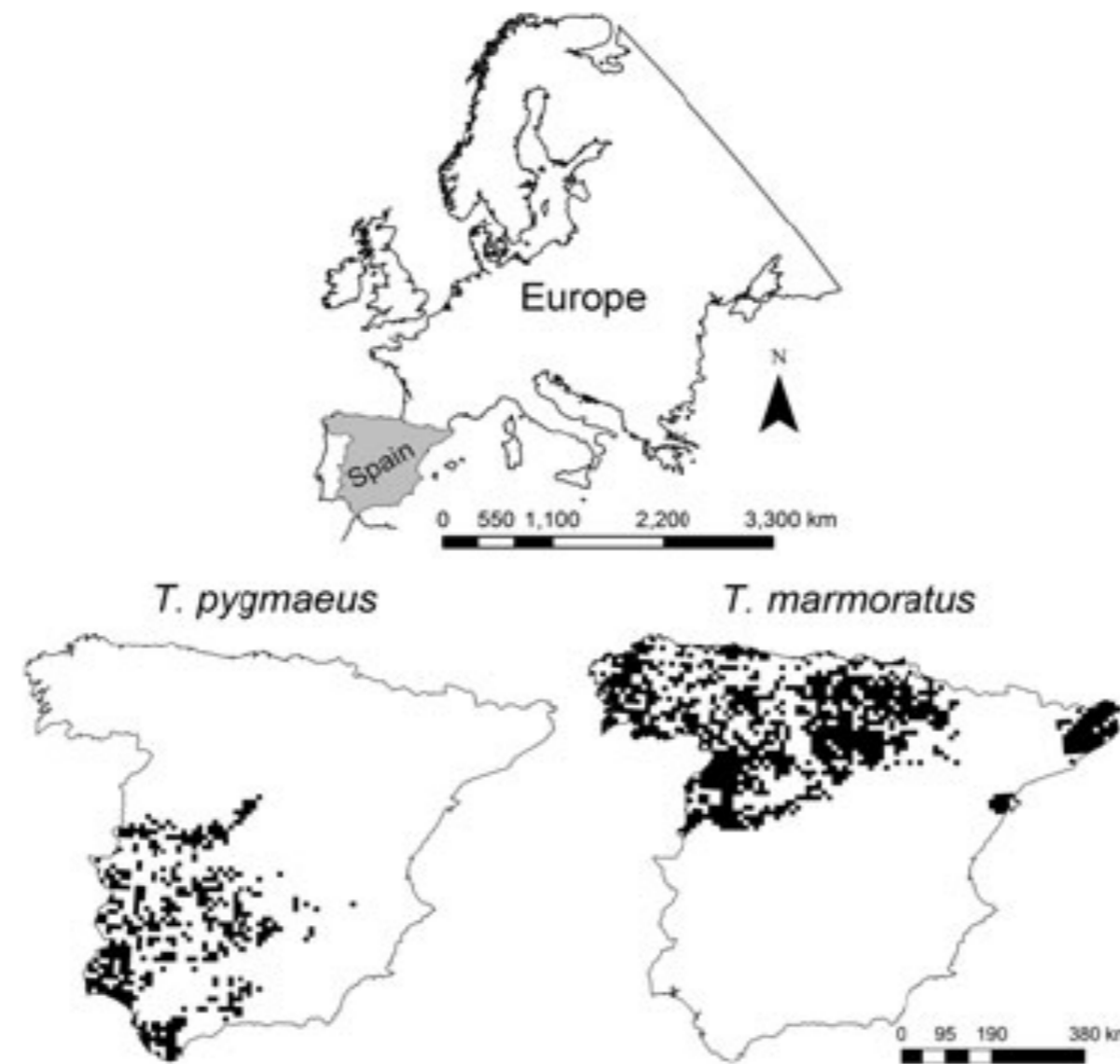
It is especially evident that the rarer taxon changes in suitability in consequence of climate changes is badly underestimated by the second approach.



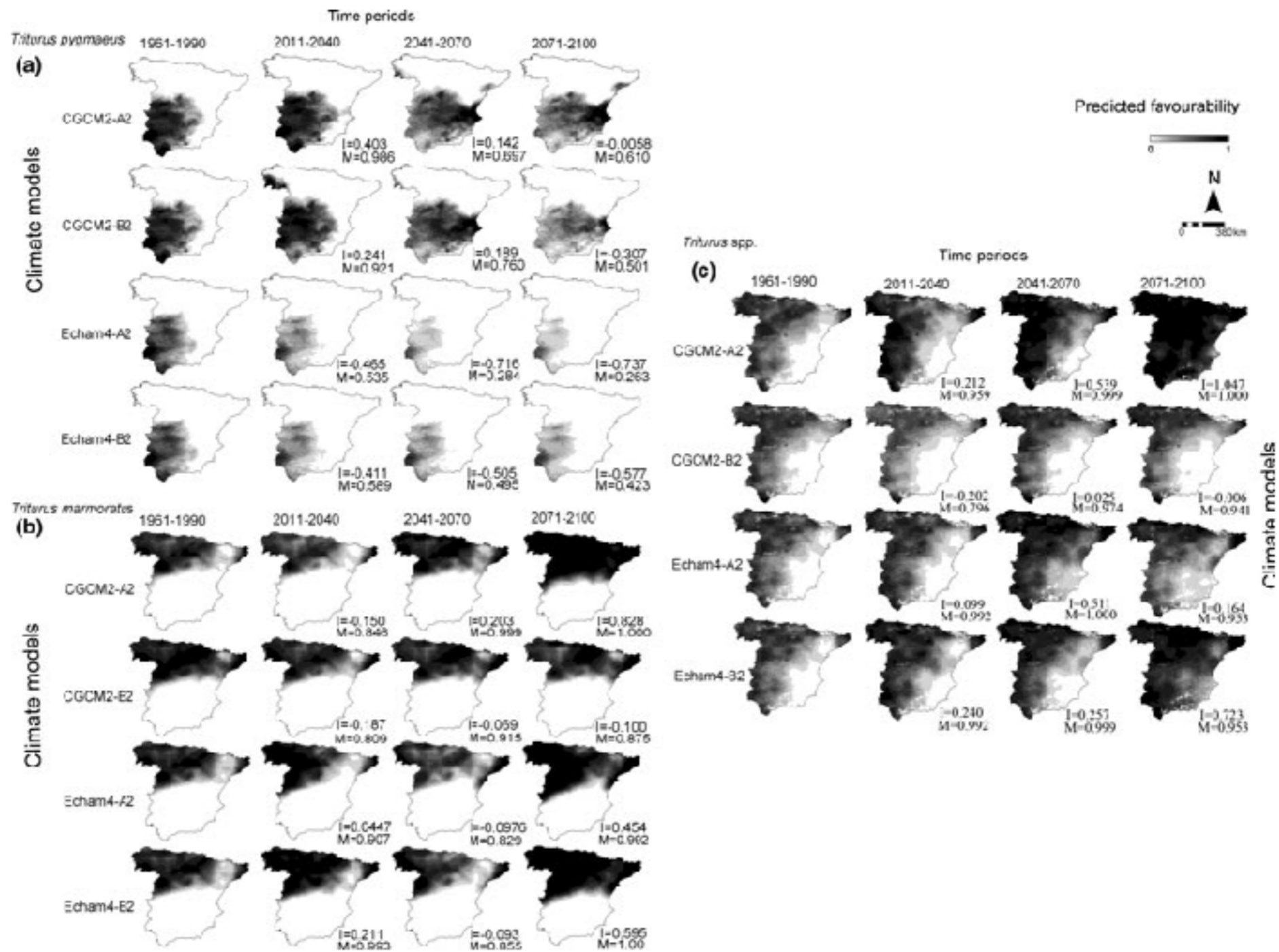
*Triturus pygmaeus*



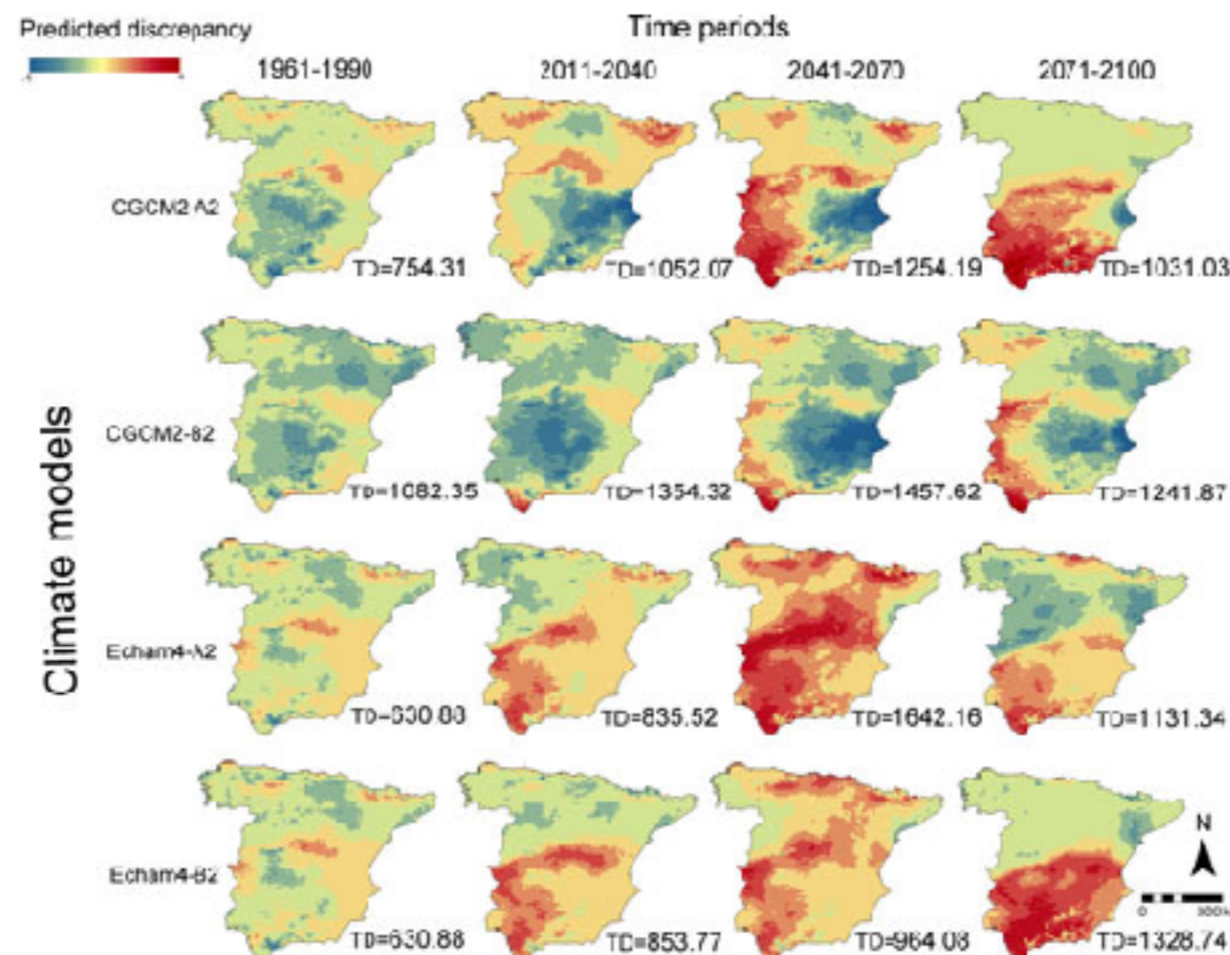
*Triturus marmoratus*



**Figure 1** The study area in the European context. Current distributions of *Triturus pygmaeus* and *T. marmoratus* represented in 10 km × 10 km UTM cells within the study area of mainland Spain (data from Pleguezuelos *et al.*, 2004).



**Figure 2** Favourability in mainland Spain predicted for (a) *Triturus pygmaeus* and (b) *T. marmoratus*, as separate species, and (c) for both *Triturus* taxa as belonging to the same species, respectively, according to each climate model and for each time period – ranging from 0 (low favourability) to 1 (high favourability). Increment (*I*) and maintenance (*M*) values are shown at the lower right corner of each map.



**Figure 3** Discrepancy values (difference between favourability predicted for both *Triturus* taxa as belonging to the same species and favourability predicted for the combination of *T. pygmaeus* and *T. marmoratus* as separate species) in mainland Spain according to each climate model and for each time period. When favourability was maximum for the *Triturus* spp. and minimum for the combination of the species, discrepancy was +1 (shown in red); in the opposite case, discrepancy was -1 (shown in blue). Total discrepancies (sum of all absolute discrepancy values, TD) are shown at the lower right corner of each map.

In this case, it is evident that an incorrect taxonomic delimitation of one or more taxa can badly influence the results of a model, together with the interpretation of the results.

This is particularly relevant when these results have a consequence in conservation practices.

Amphibians are seen as possibly the most endangered group because of human activities. Developing incorrect conservation practices following misleading models, generated by incorrect taxonomic knowledge (**taxonomic impediment**), could play a major role in the future loss of biodiversity, especially in extremely vulnerable groups.

An interesting approach to data cleaning of datasets, and in particular for datasets obtained by the GBIF, is depicted by Smith et al., 2016.

In the paper, the authors used GenBank data to develop a niche model of the lichen *Usnea longissima*. Once developed, the model has been compared with the data obtained from the GBIF for the same taxon.

The comparison evidence a huge number of incorrect records from the GBIF dataset, from ca. 600 to more than 3000, depending on the limits set in the model. In particular, it has been evidenced that all the occurrences of *U. longissima* from the southern hemisphere are incorrect, and derived from misidentifications, since the tropical area is a strong ecological barrier for the species.

There is anyway a further issue: even if an occurrence point is falling in the predicted area, it does not mean that it is a correct occurrence record. The principle of accepting records which fall inside the known distribution of a taxon can lead to overestimation of populations size, or, as an example, to an underestimation of local varieties in comparison with nominal species, or most common varieties.

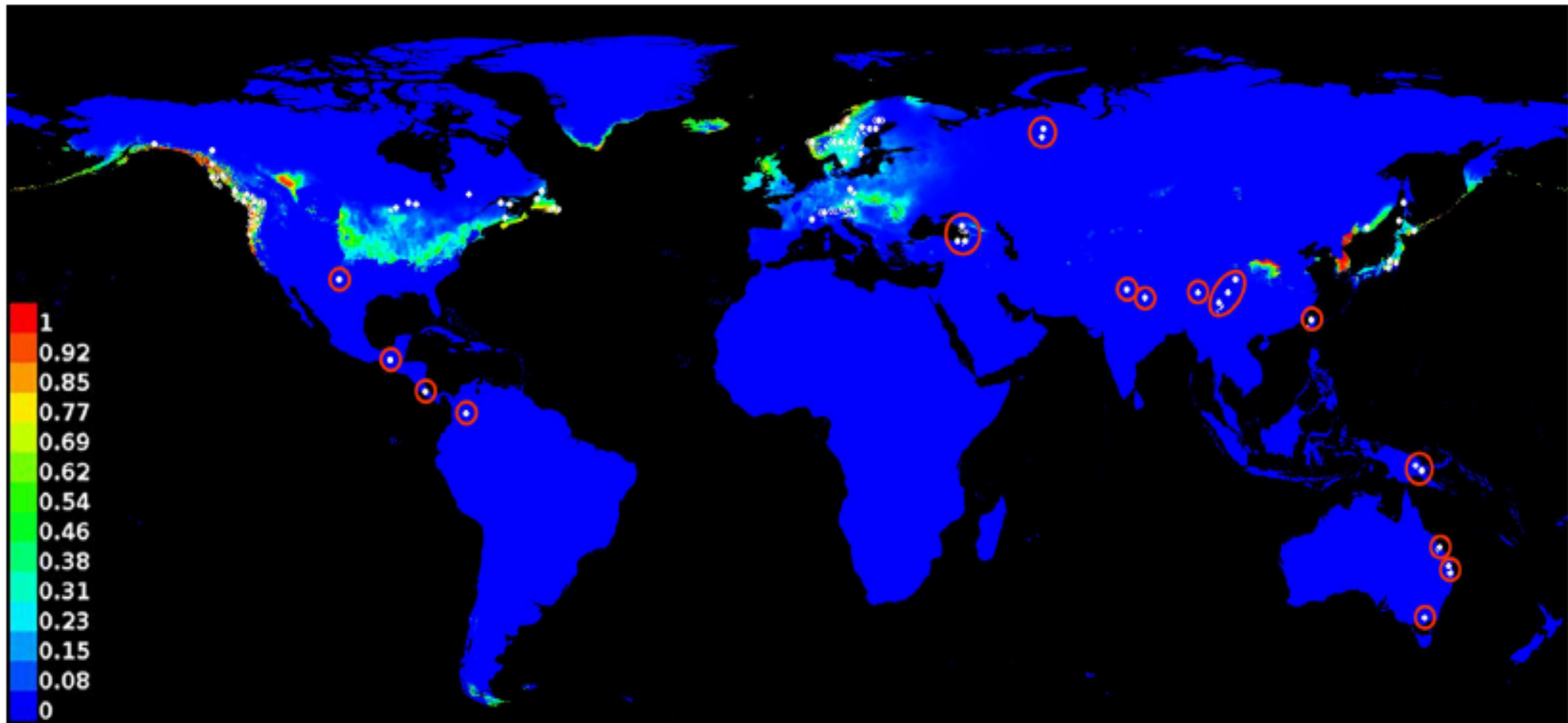


Fig 2. Best fitting MexEnt model for *Usnea longissima* based on 1477 sequenced samples corresponding to 160 localities from Rolstad et al. (2013), indicated by shaded areas ranging from pale blue-green to red. Bright blue areas indicate range of predicted absence. Map is overlaid by occurrence records from GBIF (white dots), and those falling outside the predicted range are marked with red circles. One dot may include more than one GBIF record ([S1 Table](#)).

**However....**



**...even with their limits and issues, PDB are the base on which we build SDMs.**

**Thus, what is required are tools for quality control, and a lot of work to select properly the data we aim at using.**

# **How to obtain PBD**

There are three way to obtain PBD:

1. Go in the field and collect them
2. Find and prepare an existing dataset
3. Get them from an online repository

The third is obviously the simplest, since downloading data from the GBIF is a quick operation, which could be done by a web browser, or directly in R.

However, when downloading PBD from the GBIF, one intrinsically accepts to trust them, since there is no way (but for NHM specimens) to actually falsify them. This is normally done by many researchers every day, all over the world. It can be a relevant issue, but it is often overlooked. Plus, by citing the dataset, each inference made on the data is falsifiable, thus providing a sort of “alibi”, an implicit warning sign, stating something like: *“Ehi, we obtain this, but if you want to check, you can anytime. Do not use our results without checking please”*. Problem is, these warnings are often “written” in veeeeery small characters. The consequence is that often results of many researches are accepted in scientific journals without an adequate review, and maybe used as a baseline for other researches, or worse, for relevant decision making on the management of ecosystems.

In any case, however, the GBIF - as well as other online data aggregators - remains a fundamental resource for modern research.

Dataset can be obtained from the GBIF in two ways

A) by a web browser

B) directly from R

By mean of a web browser it is simpler to set all the relevant filters, and especially do delimit the survey area, by simply drawing a polygon on a map (see later).

However, at the very end, the two processes produces the same result as far as data are concerned.

This is not true for DOI and persistence of the downloaded dataset, which could not be assured by downloading data through R.

Get data How-to Tools Community About

GBIF | Global Biodiversity Information Facility

## Free and open access to biodiversity data

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

WHAT IS GBIF? ABOUT GBIF ITALY

Occurrence records  
1.403.573.093

Datasets  
51.760

Publishing institutions  
1.587

Peer-reviewed papers using data  
4.362

News: Call for nominations to the 2020 GBIF Young Researchers Award

Data use: Relying on biodiversity science to inform art history

News: 2020 Ebbe Nielsen Challenge seeks open-data innovations for biodiversity

News: Biodiversity Summit 2020 postponed

# Obtaining PBD

Get data How-to Tools Community About

Occurrences

SEARCH OCCURRENCES | 1,403,576,826 RESULTS

Search all fields

GALLERY MAP TAXONOMY METRICS DOWNLOAD

Simple Advanced

Scientific name	Country or area	Coordinates	Month & year
<i>Peitotecus erithacus</i> Linnaeus, 1758			2020 January
<i>Asplenium radicans</i> L.	Brazil	25.2S, 50.0W	2020 January
<i>Hypoxylon cercidicola</i> (Berk. & M.A.Curtis e...	Austria	48.4N, 16.2E	2020 January
<i>Hemitrichia intorta</i> (Lister) Lister	Austria	48.2N, 16.4E	2020 January
<i>Melithreptus gularis</i> (Gould, 1837)	Australia	35.0S, 138.6E	2020 January
<i>Melithreptus gularis</i> (Gould, 1837)	Australia	35.0S, 138.6E	2020 January
<i>Aves</i>	Australia	34.5S, 135.6E	2020 January
<i>Leporilius</i> Thomas, 1906	Australia	31.2S, 141.5E	2020 January
<i>Leporilius</i> Thomas, 1906	Australia	31.0S, 125.3E	2020 January
<i>Macroderma gigas</i> (Dobson, 1880)	Australia	26.4S, 131.7E	2020 January
<i>Dasyercus cristicauda</i> (Krefft, 1867)	Australia	30.5S, 131.8E	2020 January
<i>Macroderma gigas</i> (Dobson, 1880)	Australia	26.1S, 130.1E	2020 January
<i>Macroderma gigas</i> (Dobson, 1880)	Australia	26.2S, 129.1E	2020 January

# Obtaining PBD

Get data How to Tools Community About

Occurrences

SEARCH OCCURRENCES | 481,742 RESULTS

Psittacula krameri

Did you mean  
Psittacula krameri (Scopoli, 1769) the Species?

YES NO

Simple Advanced

Scientific name	Country or area	Coordinates	Month & year
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	55.4N, 12.8E	2014 September
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	56.9N, 14.8E	2012 March
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	59.8N, 13.1E	2019 November
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	55.4N, 12.8E	2014 September
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	60.7N, 17.1E	2013 February
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	59.8N, 13.1E	
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	59.5N, 17.6E	2012 March
<i>Psittacula krameri</i> (Scopoli, 1769)	Norway	59.4N, 10.7E	2018 February
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	56.9N, 14.8E	2012 November
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	58.7N, 17.5E	
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	59.8N, 13.1E	2018 October
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	59.5N, 17.6E	2012 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Sweden	56.2N, 12.8E	2015 June

# Obtaining PBD

Get data How to Tools Community About

Occurrences

SEARCH OCCURRENCES | 338,166 RESULTS

Search all fields

Simple Advanced

Scientific name

- Psittacula krameri* (Scopoli, 1769)

Basis of record

- Observation 124
- Machine observation 33
- Human observation 336,534
- Material sample 22
- Literature 0
- Preserved specimen 1,434
- Fossil specimen 0
- Living specimen 2
- Unknown 17

Location

Year

Month

Dataset

Country or area

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Scientific name	Country or area	Coordinates	Month & year
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	51.1N, 6.9E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.6N, 8.4E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	50.8N, 7.0E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.5N, 8.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	50.0N, 8.3E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.5N, 8.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	51.2N, 6.8E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	50.0N, 8.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.6N, 8.4E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.5N, 8.4E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Portugal	38.8N, 9.1W	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Germany	49.6N, 8.4E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Portugal	38.8N, 9.1W	2020 January



# Obtaining PBD

The screenshot shows a web interface for searching species occurrences. The top navigation bar includes 'Get data', 'How-to', 'Tools', 'Community', and 'About'. The main header displays 'Occurrences' and 'SEARCH OCCURRENCES | 2,067 RESULTS'. Below the header, there are tabs for 'TABLE', 'GALLERY', 'MAP', 'TAXONOMY', 'METRICS', and 'DOWNLOAD'. The search results are displayed in a table with columns for 'Scientific name', 'Country or area', 'Coordinates', and 'Month & year'. The results list multiple occurrences of *Psittacula krameri* (Scopoli, 1769) in Italy, with coordinates and dates ranging from 2020 January to 2020 February.

The 'Location' filter panel on the left is highlighted with a red hand-drawn circle. It contains the following options:

- No preference
- Including coordinates
- Without coordinates
- Include records where coordinates are flagged as suspicious

Below the filter options is a map of Europe with a blue polygon drawn over Italy. The map includes labels for various countries and cities. At the bottom of the map, there is a checkbox and a text input field for a polygon selection tool, which is currently checked and contains the text: `POLYGON((5.625 35.17381,20.91797 35.17381,2...`

Scientific name	Country or area	Coordinates	Month & year
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	44.3N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	44.4N, 8.8E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	44.4N, 8.8E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.1N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	43.8N, 11.3E	2020 February

# Obtaining PBD

Get data How to Tools Community About

Occurrences 3

SEARCH OCCURRENCES | 1,632 RESULTS

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Scientific name	Country or area	Coordinates	Month & year
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	44.3N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.6N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.4N, 8.8E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	44.4N, 8.8E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.4N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.1E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.5N, 9.2E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	43.8N, 11.3E	2020 February

# Obtaining PBD

The screenshot shows the GBIF Occurrences interface. The top navigation bar includes 'Get data', 'How-to', 'Tools', 'Community', and 'About'. On the right, there are icons for a map, text search, a magnifying glass, and a chat bubble, along with a 'Login' button. The main content area displays a list of occurrences for *Psittacula krameri* (Scopoli, 1769) in Italy. The list includes columns for species name, country, coordinates, and date. A red hand-drawn circle highlights the 'Dataset' filter menu on the left, which lists various datasets such as 'EDD - eBird Observation Dataset' (1,062 records) and 'CEN\_PACA\_2017\_12\_18' (306 records). Below the list, there is a pagination bar with 'Previous', '1', '2', '3', '4', '5', '...', and 'Next' buttons. The footer contains links for 'What is GBIF?', 'API', 'FAQ', 'Newsletter', 'Privacy', 'Terms and agreements', 'Citation', 'Code of Conduct', and 'Acknowledgements', along with contact information for the GBIF Secretariat in Copenhagen, Denmark. Social media icons for Facebook, Twitter, LinkedIn, YouTube, and Instagram are also present.

Species	Country	Coordinates	Date
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	43.5N, 11.3E	2020 February
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 March
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2020 March
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	45.4N, 9.1E	2020 March
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2019 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	41.9N, 12.5E	2019 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	46.5N, 11.4E	2019 January
<i>Psittacula krameri</i> (Scopoli, 1769)	Italy	42.0N, 12.5E	2019 January

# Obtaining PBD

The screenshot shows the GBIF Occurrences search results page for *Psittacula krameri*. The page is divided into several sections:

- Header:** Navigation links (Get data, How-to, Tools, Community, About) and a search bar with the text "flecte".
- Search Results:** "SEARCH OCCURRENCES | 1,632 RESULTS".
- Filters:** A sidebar on the left with search fields and filters. The "Simple" filter is selected.
- Download Options:** A table of download options, highlighted with a red scribble. It includes columns for "Raw data", "Interpreted data", "Multimedia", "Coordinates", "Format", and "Estimated data size".
- Download Report:** A summary of the search results, also highlighted with a red scribble.

	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
<a href="#">SIMPLE</a>	x	✓	x	✓ (if available)	Tab-delimited CSV ⓘ	743 KB (109 KB zipped for download)
<a href="#">DARWIN CORE ARCHIVE</a>	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV ⓘ	2 MB (277 KB zipped for download)
<a href="#">SPECIES LIST</a>	x	✓	x	x	Tab-delimited CSV ⓘ	

**DOWNLOAD REPORT**

- Total: 1,632
- License: CC BY-NC 4.0
- Year range: 2010–2020
- With year: 100 %
- With coordinates: 100 %
- With taxon match: 100 %

The screenshot shows the GBIF Occurrences search interface. A modal dialog box is open in the center, titled "Free of cost – not free of responsibilities". The dialog contains the following text:

**Free of cost – not free of responsibilities**

While data from GBIF.org is free and open, please remember that by downloading this data, you are agreeing:

- to abide by the [GBIF user agreement](#)
- and, if you use the data, to [cite it appropriately](#)

Please make sure your citation includes the unique DOI (shown on the page once it refreshes). The use of properly formatted data citations ensures scientific transparency and reproducibility and enables proper attribution of credit to the data providers.

If you are analysing the data you will download, please consider referencing this citation in your Materials and methods section.

Buttons: Cancel, UNDERSTOOD

The background interface shows a search for "flechte" with 1,632 results. The left sidebar includes filters for License, Scientific name (Psittacula krameri), Basis of record, Location (including coordinates), Year (2010-2020), Month, Dataset, Country or area, Continent, Issues and flags, and Media type. The right sidebar shows download options for Tab-delimited CSV files with estimated data sizes.

Format	Estimated data size
Tab-delimited CSV	743 KB (109 KB zipped for download)
Tab-delimited CSV	2 MB (277 KB zipped for download)
Tab-delimited CSV	

DOWNLOAD REPORT

- Total: 1.632
- License: CC BY-NC 4.0
- Year range: 2010–2020
- With year: 100 %
- With coordinates: 100 %
- With taxon match: 100 %

[Get data](#)
[How-to](#)
[Tools](#)
[Community](#)
[About](#)

[Rechte](#)

DOWNLOAD | 6 APRIL 2020  
**Under processing**  
 DOI: [10.15468/dl.d3cqr6](https://doi.org/10.15468/dl.d3cqr6)

Preparing CANCEL

FILTER APPLIED 6 APRIL 2020

The download has been started and is currently being processed.  
 Please expect up to 3 hours for the download to complete. Most downloads will complete within 15 minutes.  
 A notification email with a link to download the results will be sent to the following address once ready: [martelst@units.it](mailto:martelst@units.it)

**Citation:** GBIF.org (06 April 2020) GBIF Occurrence Download <https://doi.org/10.15468/dl.d3cqr6>  
**License:** Unspecified  
 Make sure to read the [data user agreement](#) and [citation guidelines](#).

**And** API

- Year** Between start of 2010 and end of 2020
- Geometry** POLYGON((5.625 35.17381,20.91797 35.17381,20.91797 47.15984,5.625 47.15984,5.625 35.17381))
- Scientific name** *Psittacula krameri* (Scopoli, 1769)
- Has geospatial issue** false

The screenshot shows a GBIF data download page. At the top, there is a green navigation bar with links for 'Get data', 'How-to', 'Tools', 'Community', and 'About'. On the right side of this bar are icons for a heart, a document with a checkmark, a magnifying glass, and a speech bubble, along with a search box containing the text 'flechte'. Below the navigation bar, the DOI '10.15468/dl.d3cqr6' is displayed. A prominent red hand-drawn circle highlights a green 'DOWNLOAD' button in the top right corner. Below this, the text 'RERUN QUERY' is visible. The main content area has a filter applied: 'FILTER APPLIED 6 APRIL 2020'. It contains the following information: 'Citation: GBIF.org (06 April 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.d3cqr6', 'License: CC BY-NC 4.0', 'File: 90 KB Simple', and 'Involved datasets: 21'. A note states: 'Make sure to read the data user agreement and citation guidelines.' Below this, a warning message reads: 'Unless GBIF discovers citations of this download, the data file is eligible for deletion after October 6, 2020. Read more about our deletion policy.' A red hand-drawn oval highlights a row of three buttons: 'TELL US ABOUT USAGE', 'POSTPONE DELETION', and 'DELETE DOWNLOAD'. At the bottom, a filter tree is shown under the heading 'And', with the following criteria: 'Year' (Between start of 2010 and end of 2020), 'Geometry' (POLYGON((5.625 35.17381,20.91797 35.17381,20.91797 47.15984,5.625 47.15984,5.625 35.17381))), 'Scientific name' (Psittacula krameri (Scopoli, 1769)), and 'Has geospatial issue' (false). An 'API' link is located to the right of the filter tree.

downloads@gbif.org

09:24



Your GBIF data download is ready

A: Stefano Martellos

Hello flechte,

Your download is available at the following address:

<http://api.gbif.org/v1/occurrence/download/request0035948-200221144449610.zip>

When using this dataset please use the following citation:

[GBIF.org](https://doi.org/10.15468/cl.d3cqr6) (6 April 2020) GBIF Occurrence Download <https://doi.org/10.15468/cl.d3cqr6>

Download Information:

DOI: <https://doi.org/10.15468/dl.d3cqr6> (may take some hours before being active)

Creation Date: 07:23:27 6 April 2020

Records included: 1632 records from 21 published datasets

Compressed data size: 89.7 kB

Download format: simple tab-separated values (TSV)

Filter used:

Year: 2010-2020

Geometry: POLYGON((5.625 35.17381,20.91797 35.17381,20.91797 47.15984,5.625 47.15984,5.625 35.17381))

TaxonKey: Psittacula krameri (Scopoli, 1769)

HasGeospatialIssue: false

Download file retention:

Information about this download will always be available at <https://doi.org/10.15468/dl.d3cqr6> and <https://www.gbif.org/occurrence/download/0035948-200221144449610>

The simple tab-separated values (TSV) file will be kept for six months (until 6 October 2020). You can ask

us to keep the file for longer from <https://www.gbif.org/occurrence/download/0035948-200221144449610>

If you cite this download using the DOI, we will usually detect this and keep the file indefinitely.

For more information on this, see <https://www.gbif.org/faq?question=for-how-long-will-does-gbif-store-downloads>

For help with opening downloaded files, see

<https://www.gbif.org/faq?question=opening-gbif-csv-in-excel>

or the FAQ section of the GBIF website:

<https://www.gbif.org/faq>

Link to the file

Citation

Recap of query parameters

File retention



# Obtaining PBD

gbifID	datasetKey	occurrenceID	kingdom	phylum	class	order	family	genus	species				
2580194957	50c9509d-22c7-4a22-a47d-8c48425e4a7	<a href="https://www.inaturalist.org/observations/39944292">https://www.inaturalist.org/observations/39944292</a>	Animalia	Chordata	Aves	Psittaciformes	Psittacidae	Psittacula	Psittacula krameri				
2580125218	50c9509d-22c7-4a22-a47d-8c48425e4a7	<a href="https://www.inaturalist.org/observations/39969969">https://www.inaturalist.org/observations/39969969</a>	Animalia	Chordata	Aves	Psittaciformes	Psittacidae	Psittacula	Psittacula krameri				
2580	infraspecificEpithet	taxonRank	scientificName	verbatimScientificName	verbatimScientificNameAuthorship	countryCode	locality	stateProvince					
2579		SPECIES	Psittacula krameri (Scopoli, 1769)	Psittacula krameri		IT		Lombardia					
2578		SPECIES	Psittacula krameri (Scopoli, 1769)	Psittacula krameri		IT		Lazio					
2578	occurrenceStatus	individualCount	publishingOrgKey	decimalLatitude	decimalLongitude	coordinateUncertaintyInMeters	coordinatePrecision	elevation	elevationAccuracy				
2576			28eb1a3f-1c15-4a95-931a-4a90ecb574d	45.426071	9.138852	1379.0							
2574			28eb1a3f-1c15-4a95-931a-4a90ecb574d	41.852145	12.50296	17.0							
2574	depth	depthAccuracy	eventDate	day	month	year	taxonKey	speciesKey	basisOfRecord	institutionCode	collectionCode	catalogNumber	recordNumber
2573			2020-03-13T16:55:00Z	13	3	2020	2479226	2479226	HUMAN OBSERVATION	iNaturalist	Observations	39944292	
2573			2020-03-14T16:19:03Z	14	3	2020	2479226	2479226	HUMAN OBSERVATION	iNaturalist	Observations	39969969	
2573	identifiedBy	dateIdentified	license	rightsHolder	recordedBy	typeStatus	establishmentMeans	lastInterpreted					
2563		2020-03-13T23:14:44Z	CC_BY_4_0	Alexandros Quartarone	Alexandros Quartarone			2020-03-30T20:24:53.393Z					
2563		2020-03-14T16:18:20Z	CC_BY_NC_4_0	marco caporioni	marco caporioni			2020-03-30T20:24:11.745Z					
2563	typeStatus	establishmentMeans	lastInterpreted	mediaType	issue								
2563			2020-03-30T20:24:53.393Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2563			2020-03-30T20:24:11.745Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2557			2020-03-30T20:24:32.967Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2557			2020-03-30T20:24:19.080Z		COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2557			2020-03-30T20:24:59.490Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2557			2020-03-30T20:25:23.463Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2550			2020-03-30T20:24:12.286Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2549			2020-03-30T20:25:03.081Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2543			2020-03-30T20:25:04.276Z		COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2542			2020-03-30T20:24:52.022Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2542			2020-03-30T20:24:19.970Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2540			2020-03-30T20:24:04.612Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
2540			2020-03-30T20:23:56.862Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:25:31.788Z		COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:09.983Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:25:37.153Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:23:57.686Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:25:02.706Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:14.991Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:12.862Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:48.424Z	StillImage	GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:25:37.683Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:03.459Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:52.456Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:25:30.342Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:23:59.470Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								
			2020-03-30T20:24:40.112Z	StillImage	COORDINATE_ROUNDED;GEODETTIC_DATUM_ASSUMED_WGS84								

Direct download from R is possible by mean of several different functions, included in different packages.

We will use the function *gbif* of the library *dismo*

We will also perform some data cleaning, and especially:

- A) remove records with no data in the latitude or longitude fields
- B) remove duplicate records

We will also store a *comma separated value* (csv) file in the working directory, containing the data which are normally necessary to run an algorithm, i.e. longitude and latitude

**Let's switch to R**