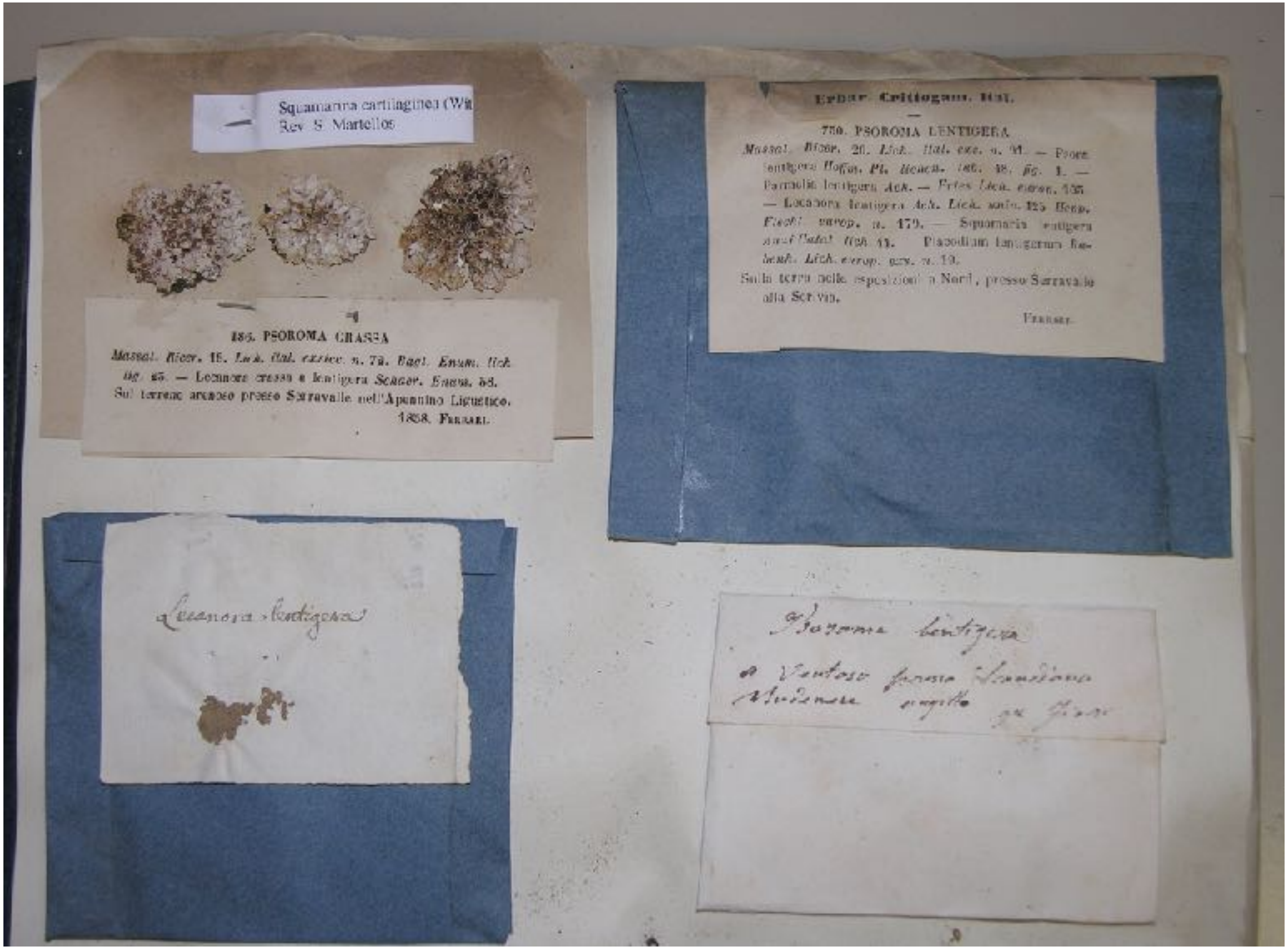
Biological data:
Primary biodiversity data

What primary biodiversity data are?

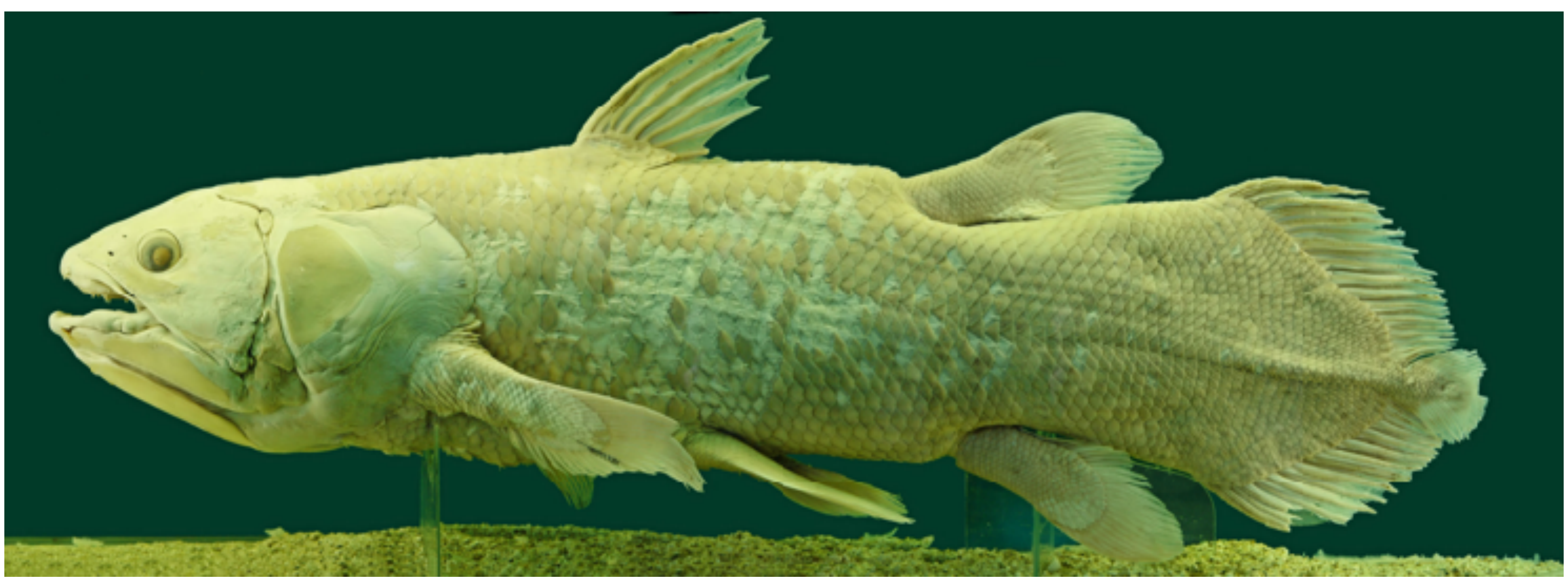
Where do they come from?

Are they complex?

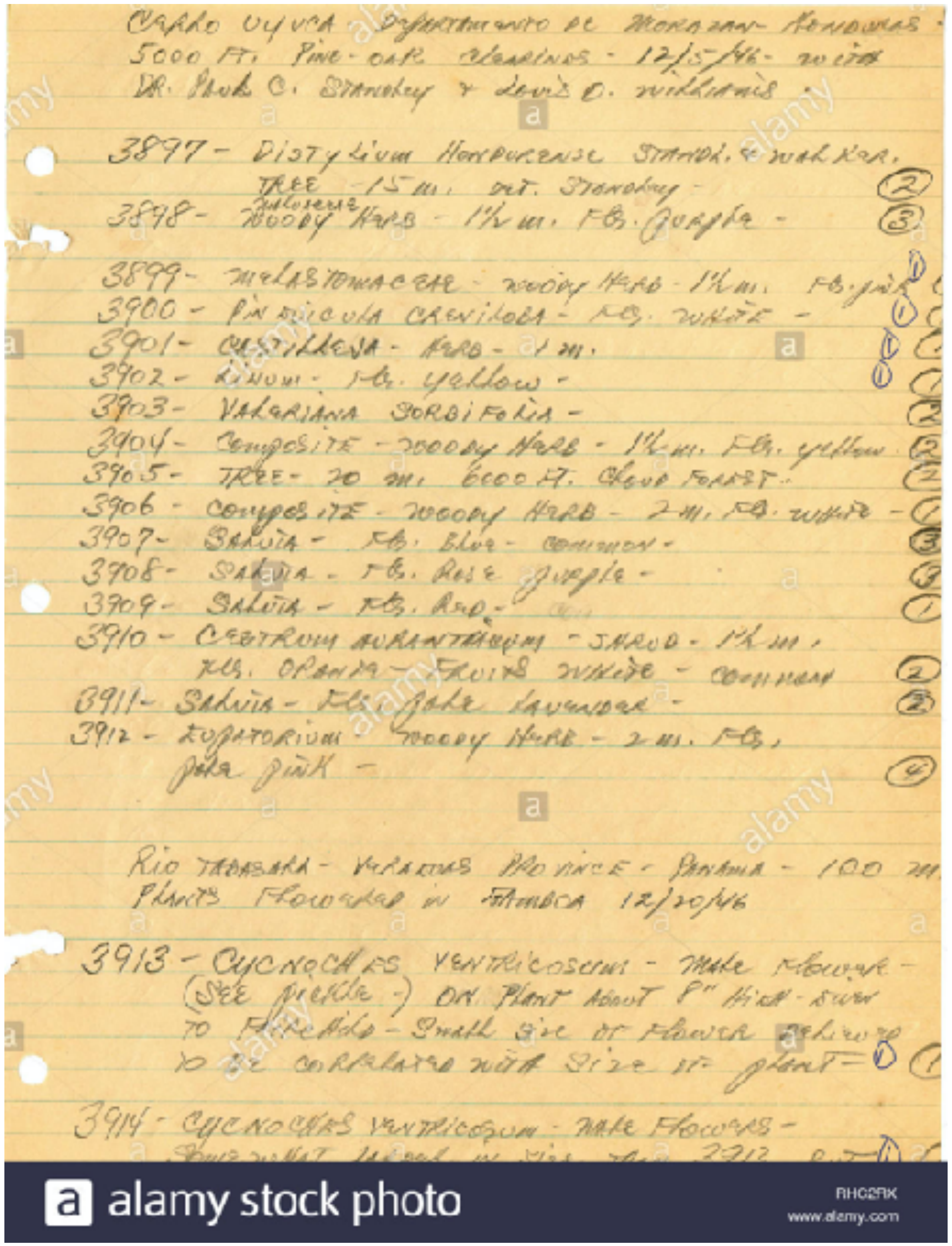
Primary biodiversity data

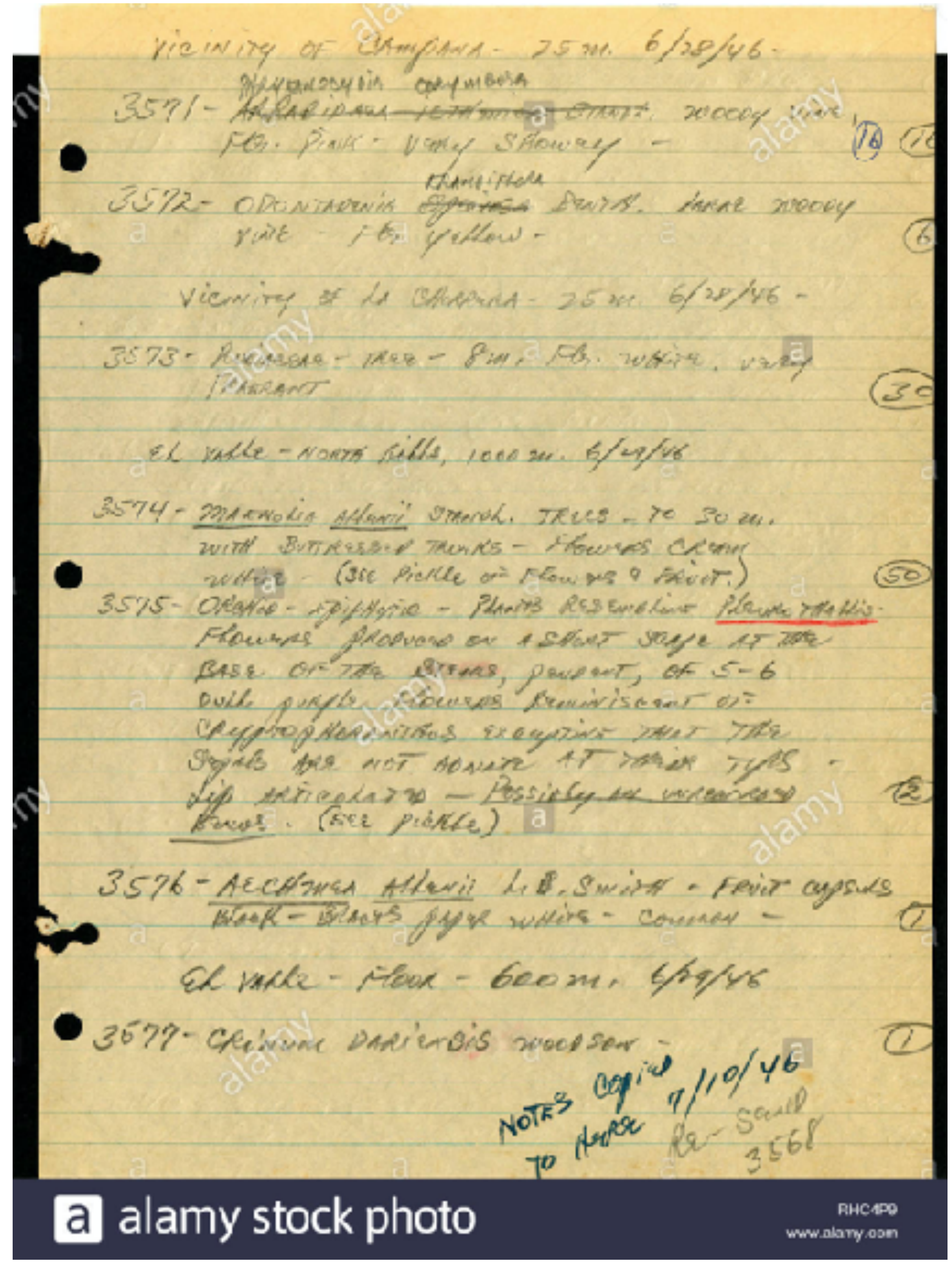


Primary biodiversity data



Primary biodiversity data





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 a 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 then 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…..

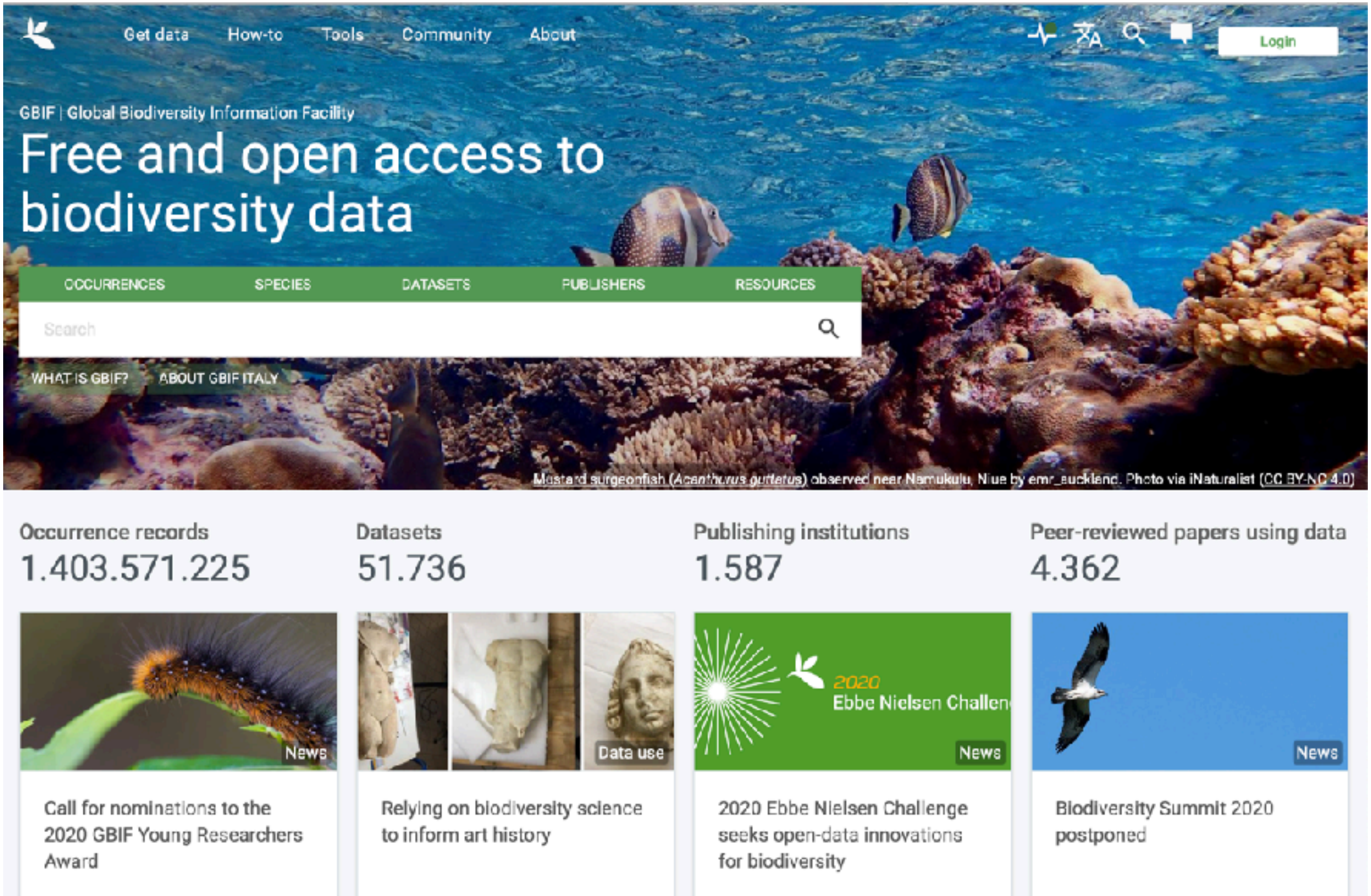
Plus, even if me and the one ho verifies the observation observe the same organism, was it the same individual?

Big problem:

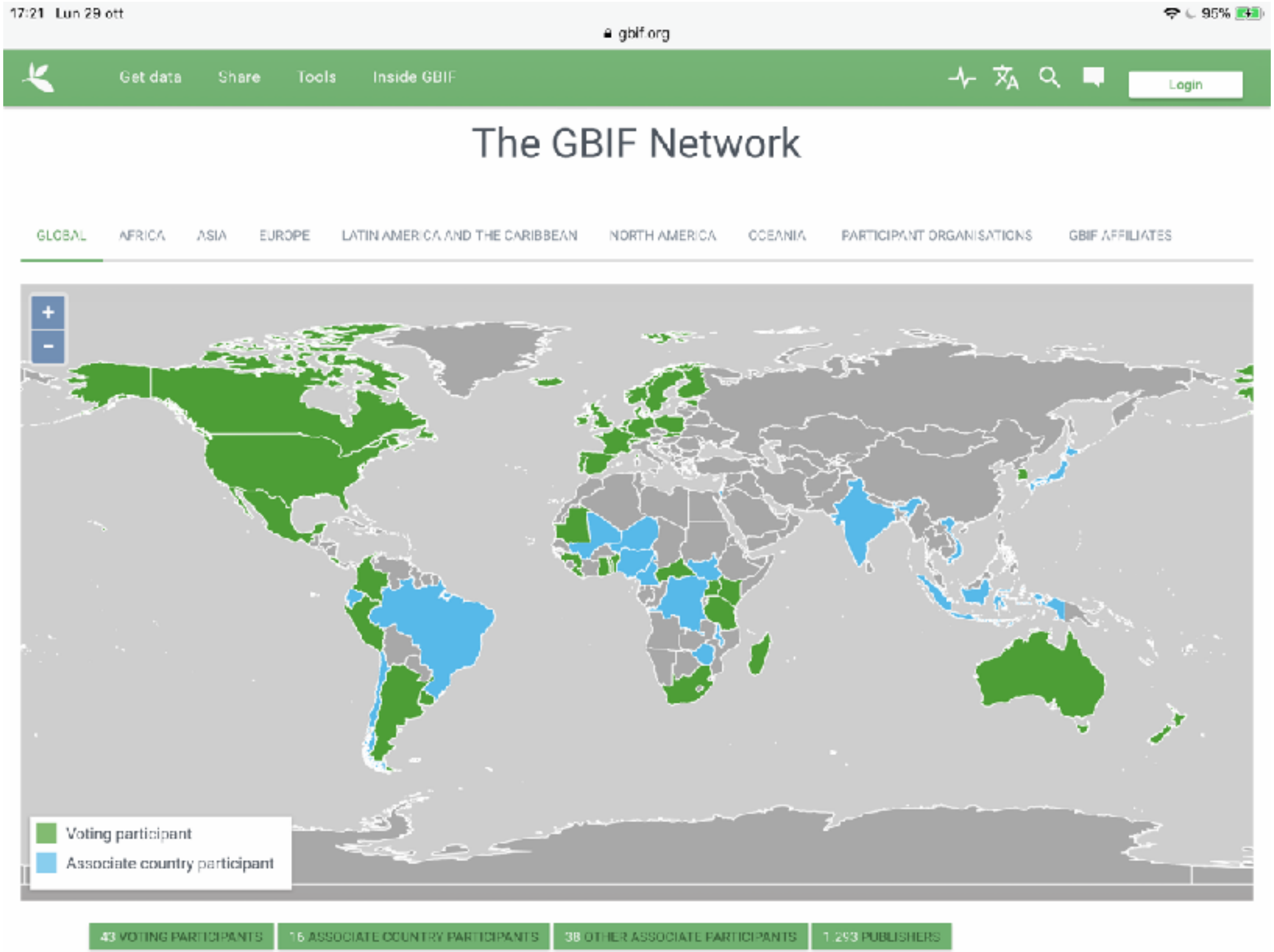
To be used, data should exist in the digital domain….

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

Primary biodiversity data: The GBIF



Primary biodiversity data: The GBIF



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

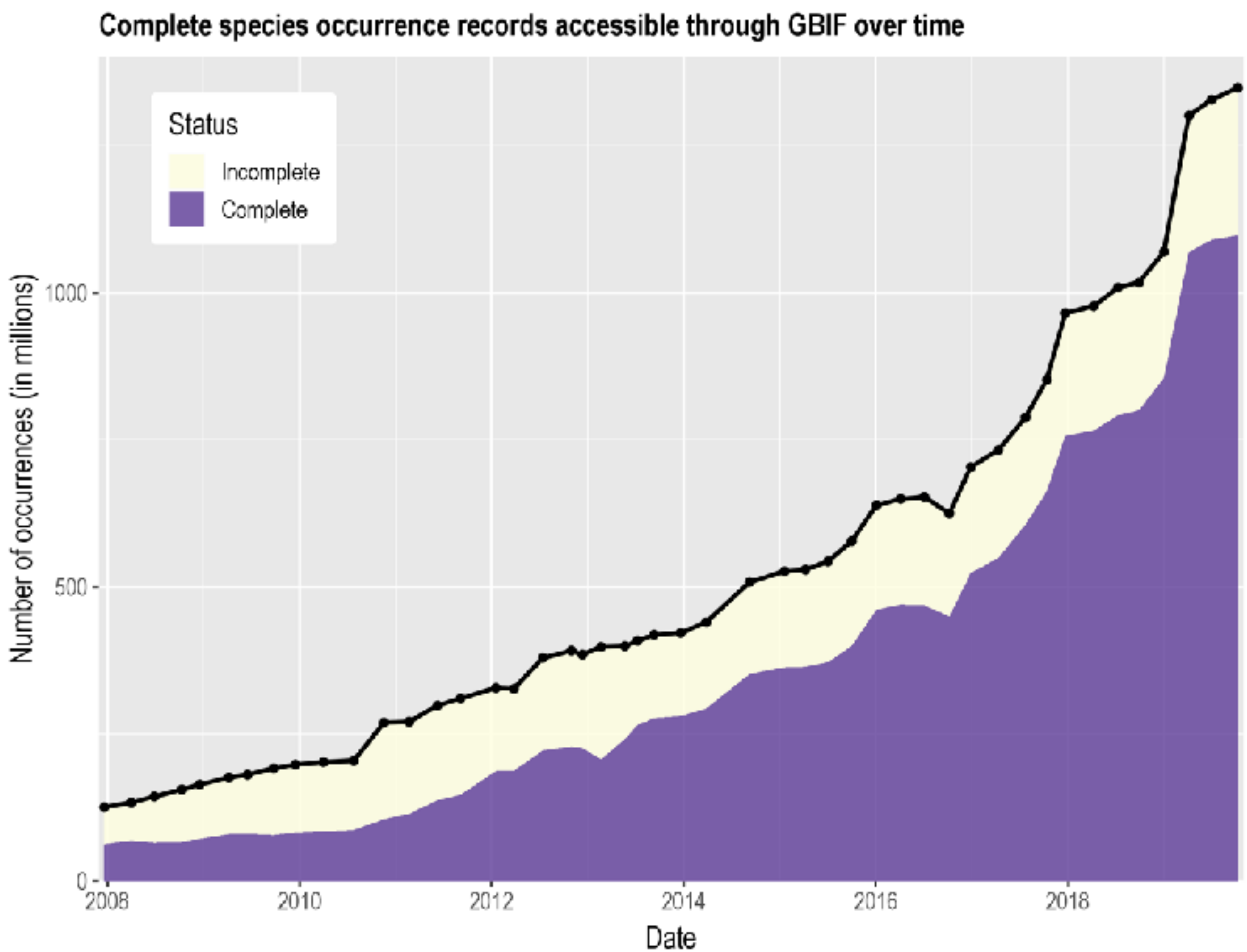
1.662.149.688 occurrence records

57.008 datasets

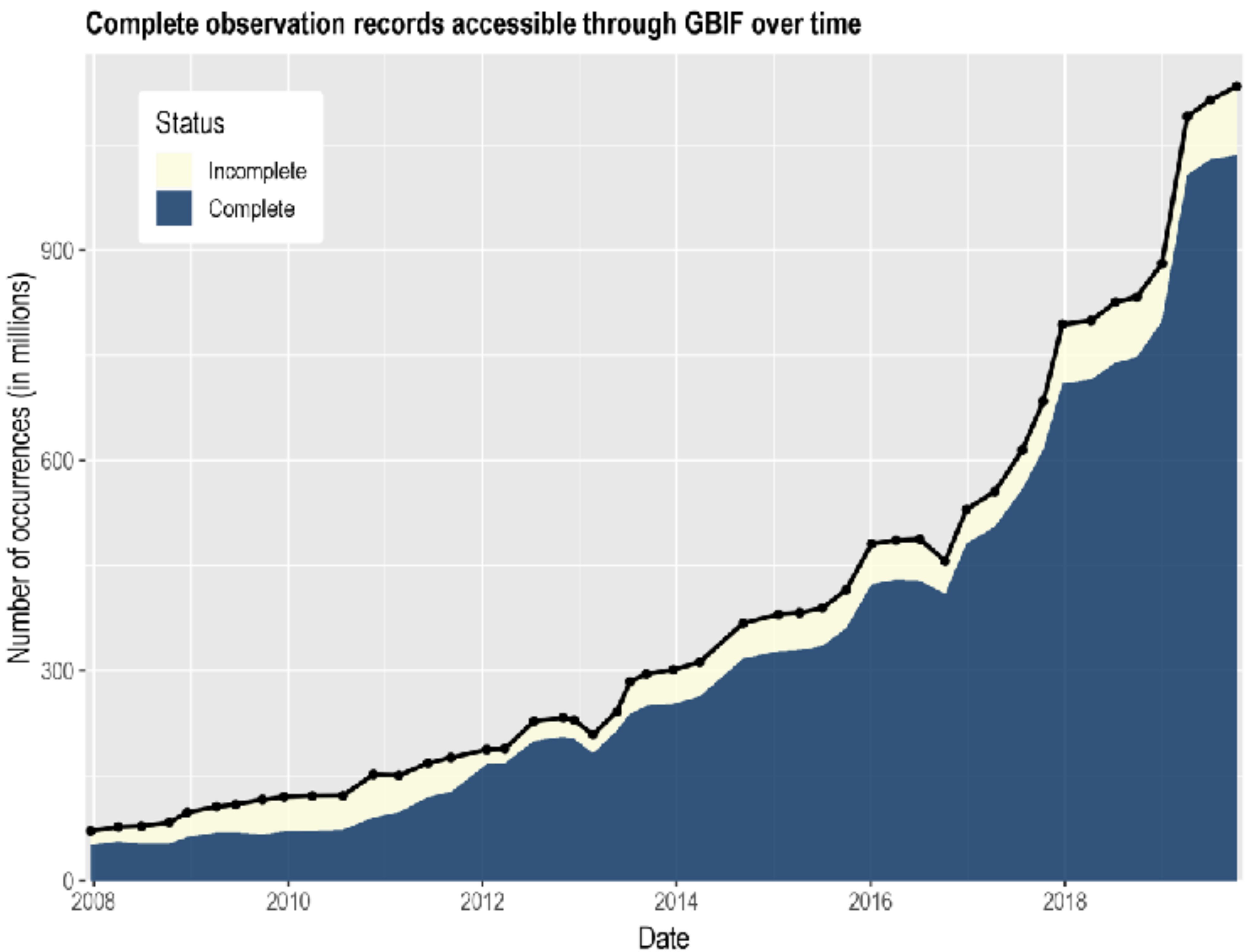
1.655 publishing institutions

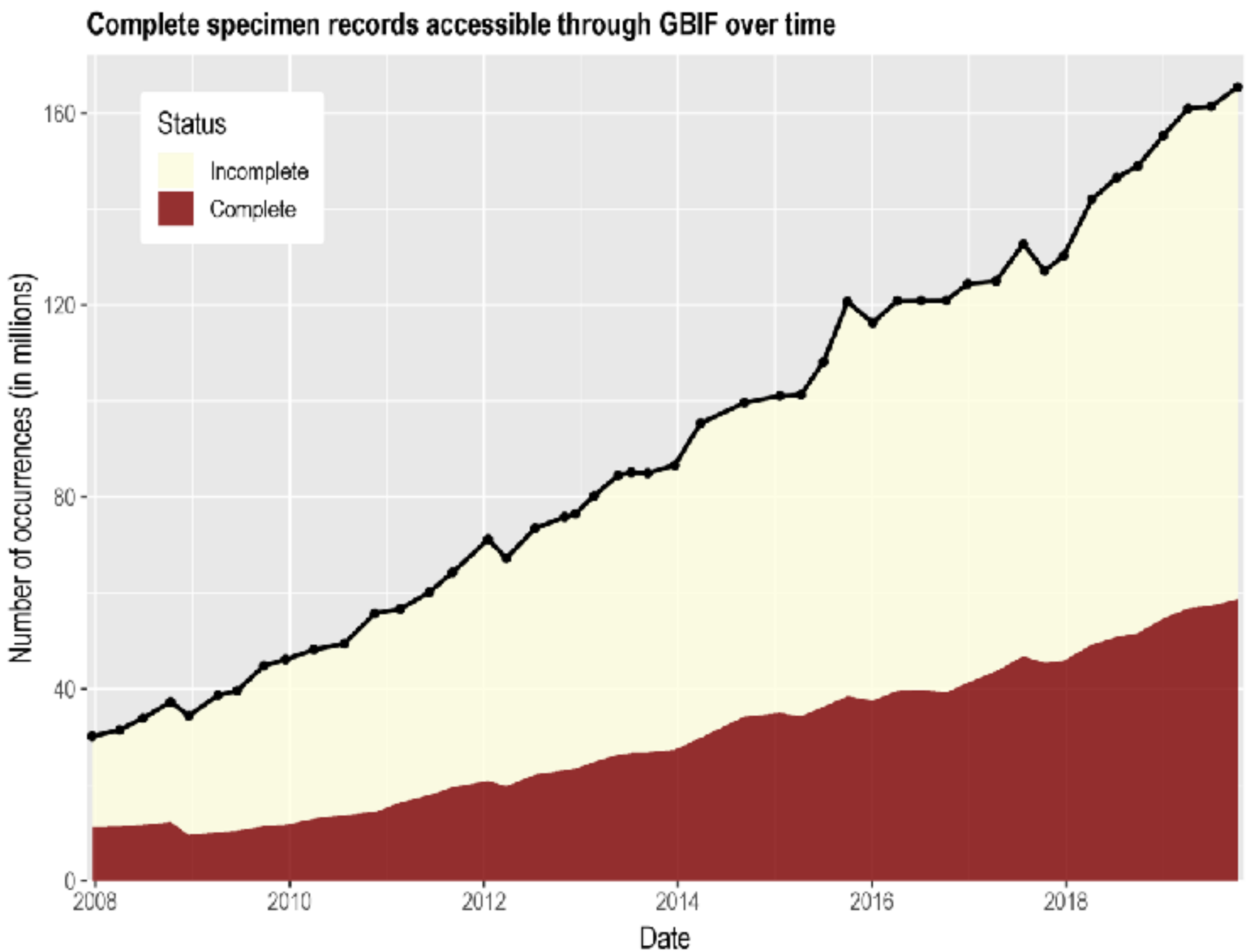
5.579 peer reviewed papers using GBIF data

Primary biodiversity data: The GBIF



Primary biodiversity data: The GBIF



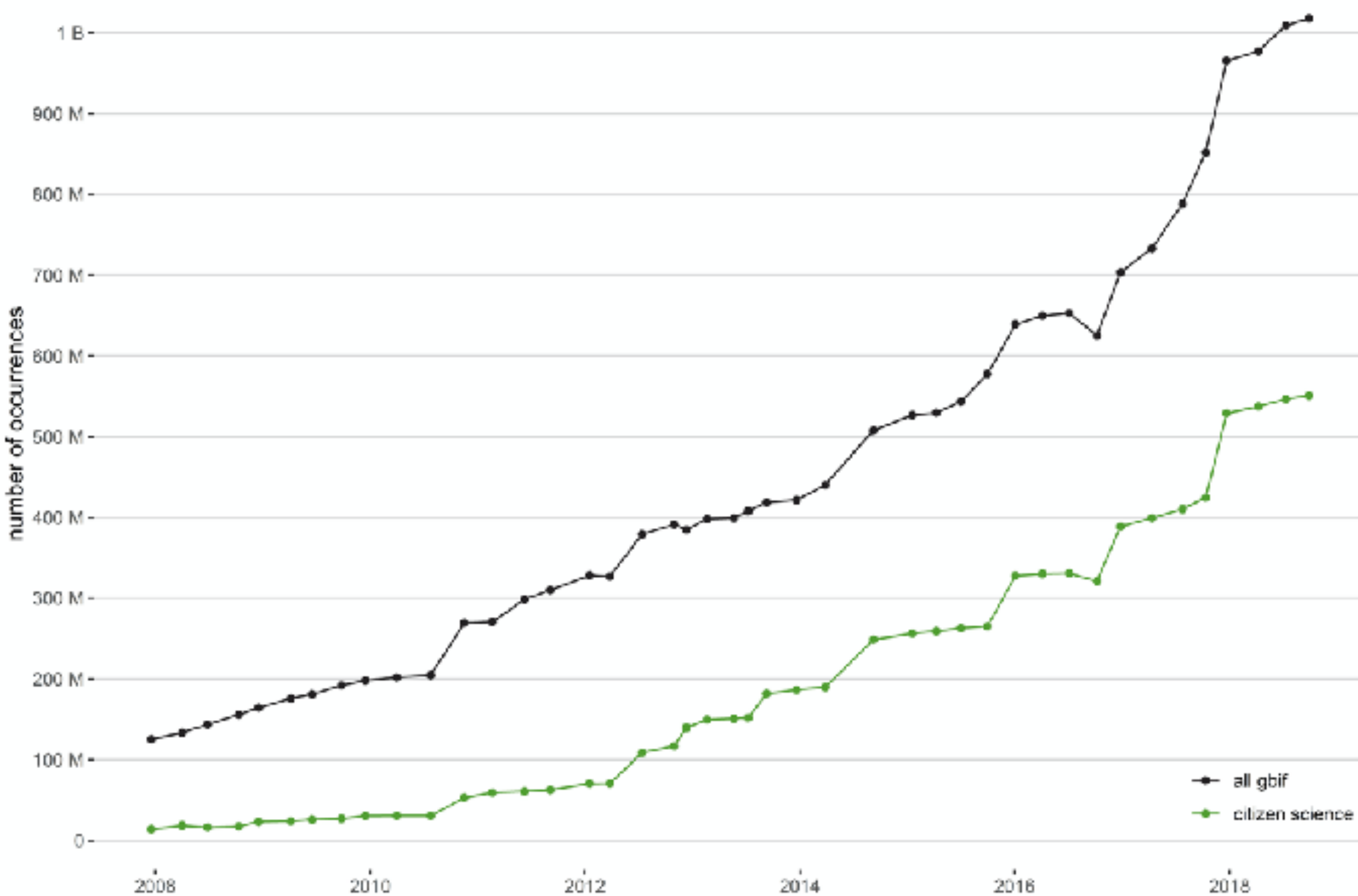

Complete specimen records accessible through GBIF over time
Status
Incomplete
Complete
Number of occurrences (in millions)
0
40
80
120
160
2008
2010
2012
2014
2016
2018
Date

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	
Observation	19.653.975
Machine observation	11.912.825
Human observation	1.105.925.840
Material sample	25.326.297
Literature	219.987
Preserved specimen	166.087.620
Fossil specimen	9.029.568
Living specimen	1.637.949
Unknown	17.147.671

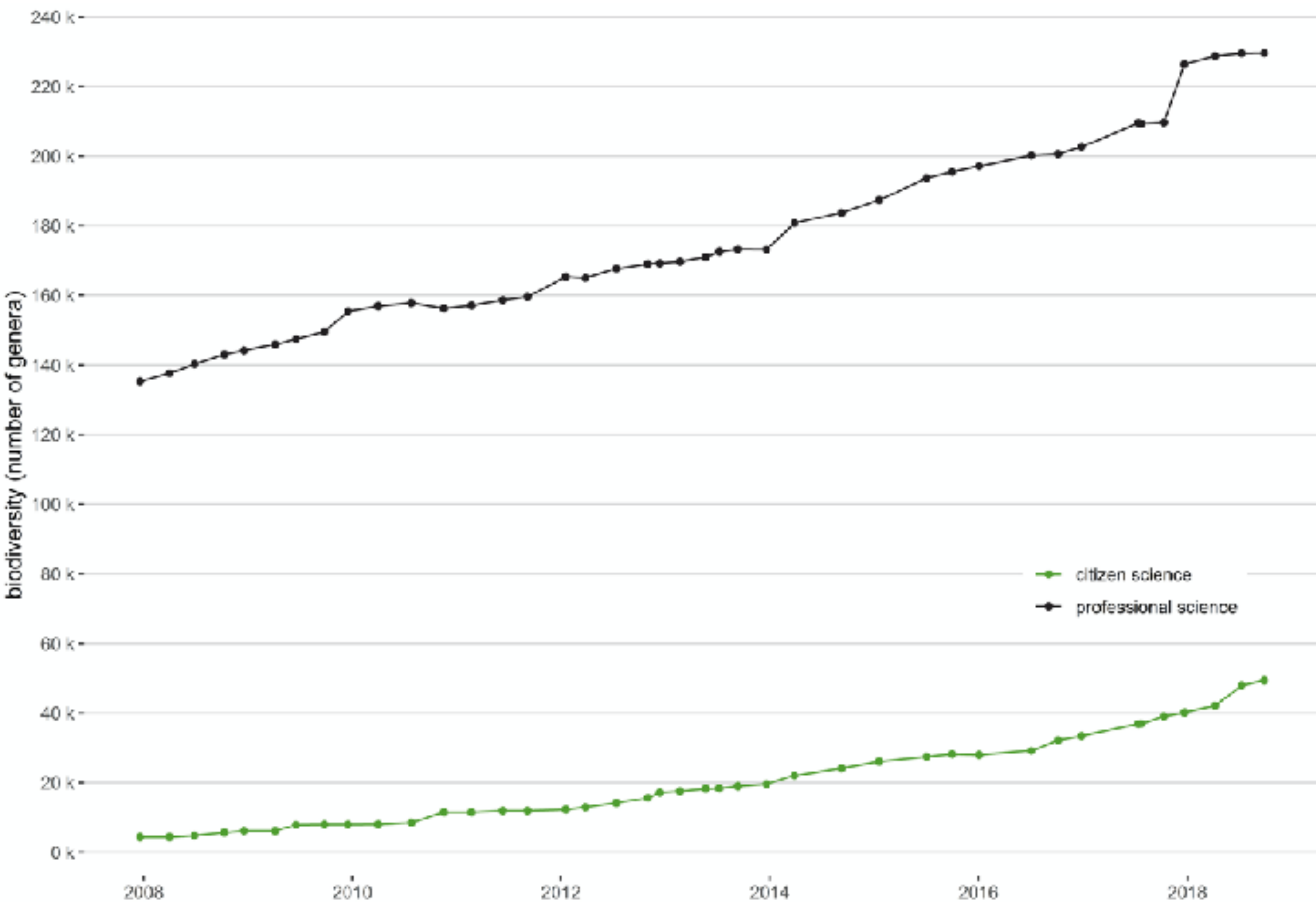
Furthermore, when they are not, their digitization is often far easier that for natural history collection specimens.

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



50% of occurrence records on GBIF are citizen science observations

However….


240 k
220 k
200 k
180 k
160 k
140 k
120 k
100 k
80 k
60 k
40 k
20 k
0 k
biodiversity (number of genera)
2008
2010
2012
2014
2016
2018
citizen science
professional science

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.

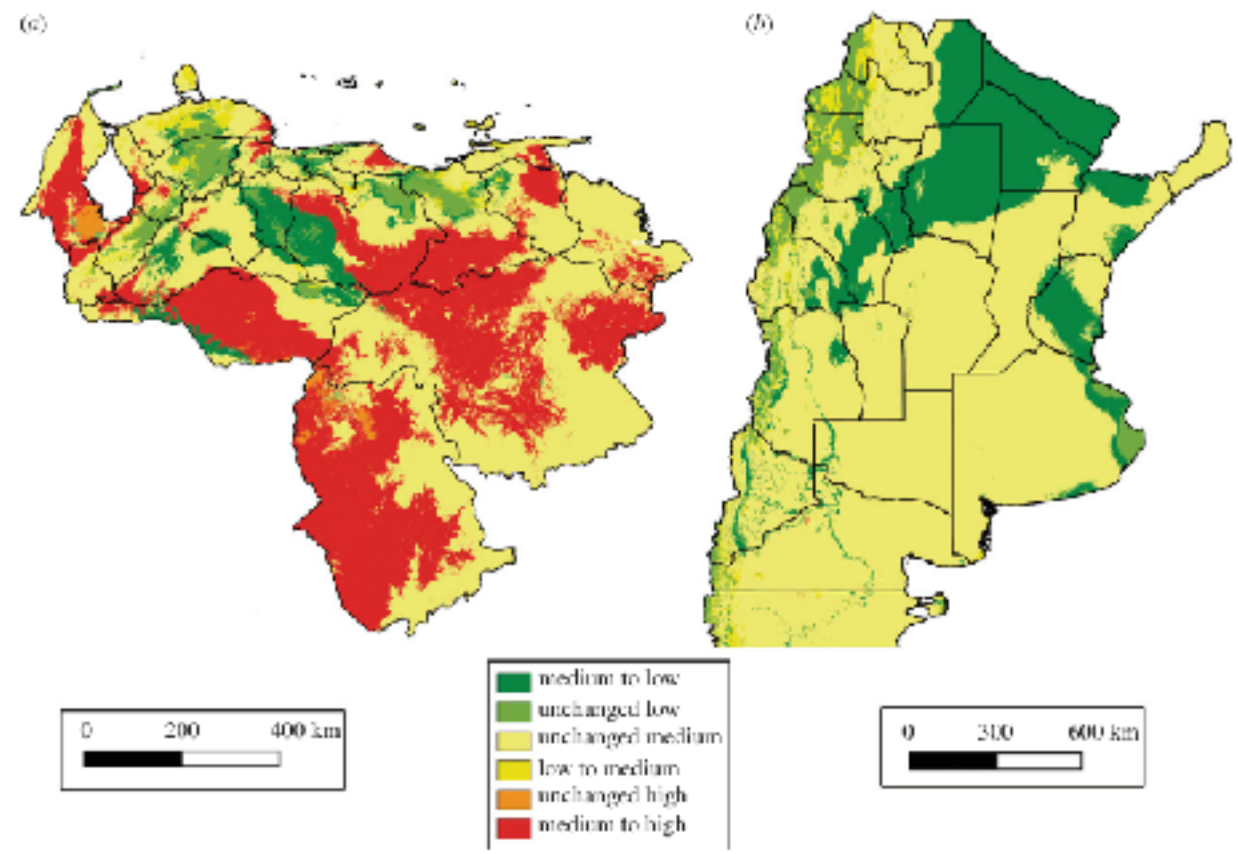


Figure 1. Climatic suitability changes from current to 2050 conditions for (a) *R. prolixus* and (b) *T. infestans*. Colours indicate climatic suitability transitions between the three main suitability categories (low, medium and high). (Online version in colour.)

PBD: poor quality issues

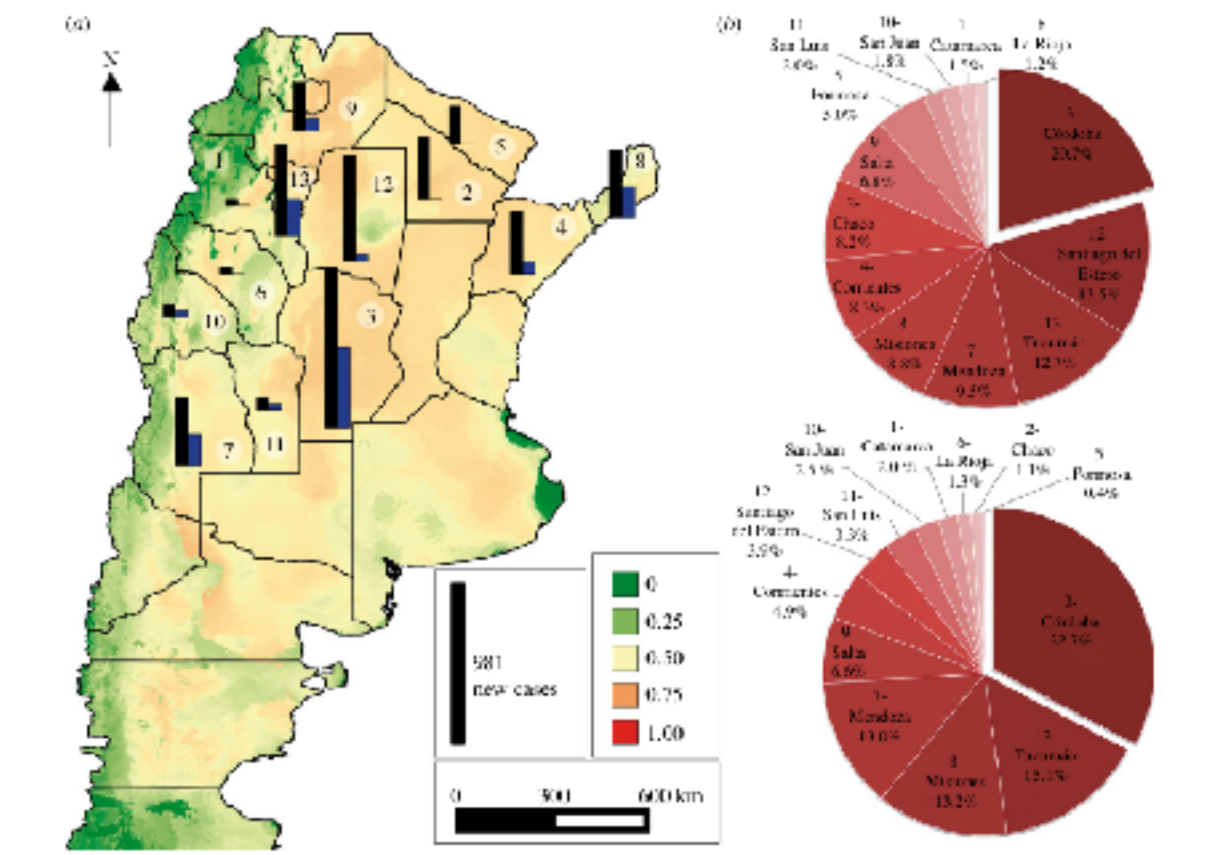


Figure 3. Current climatic suitability (base map green to red/grey to black colours) and predicted number of new human cases of *T. cruzi* infection (current: black bar, 2050: blue/grey bar) in Argentina. Provinces were numerated (white circle on the map) to facilitate the identification on the pie charts (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.)

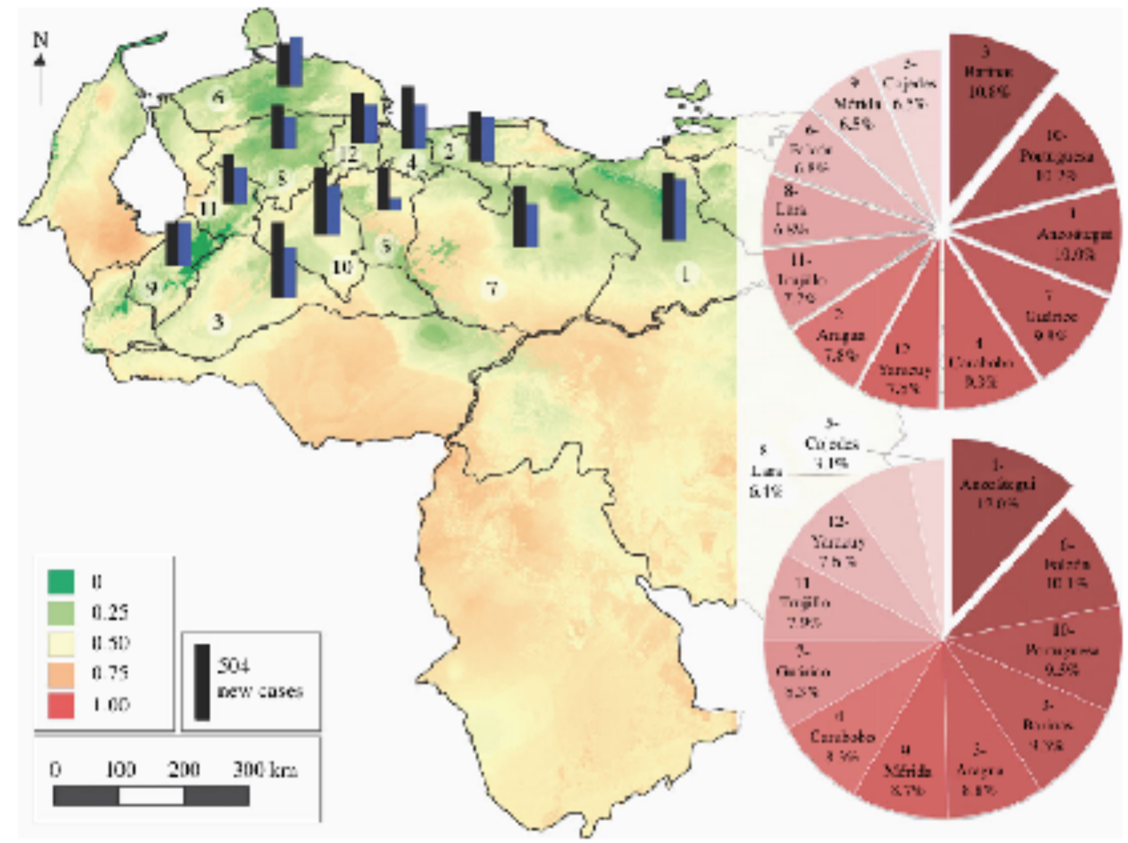


Figure 7. Current climatic suitability (base map green to red/grey to black colours) and predicted number of new cases of *T. cruzi* human infection (current: black bar, 2050: blue/grey bar) in Venezuela. States were numerated (white circle on the map) to facilitate the identification on the pie charts (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.

PBD: poor quality issues

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.

PBD: poor quality issues

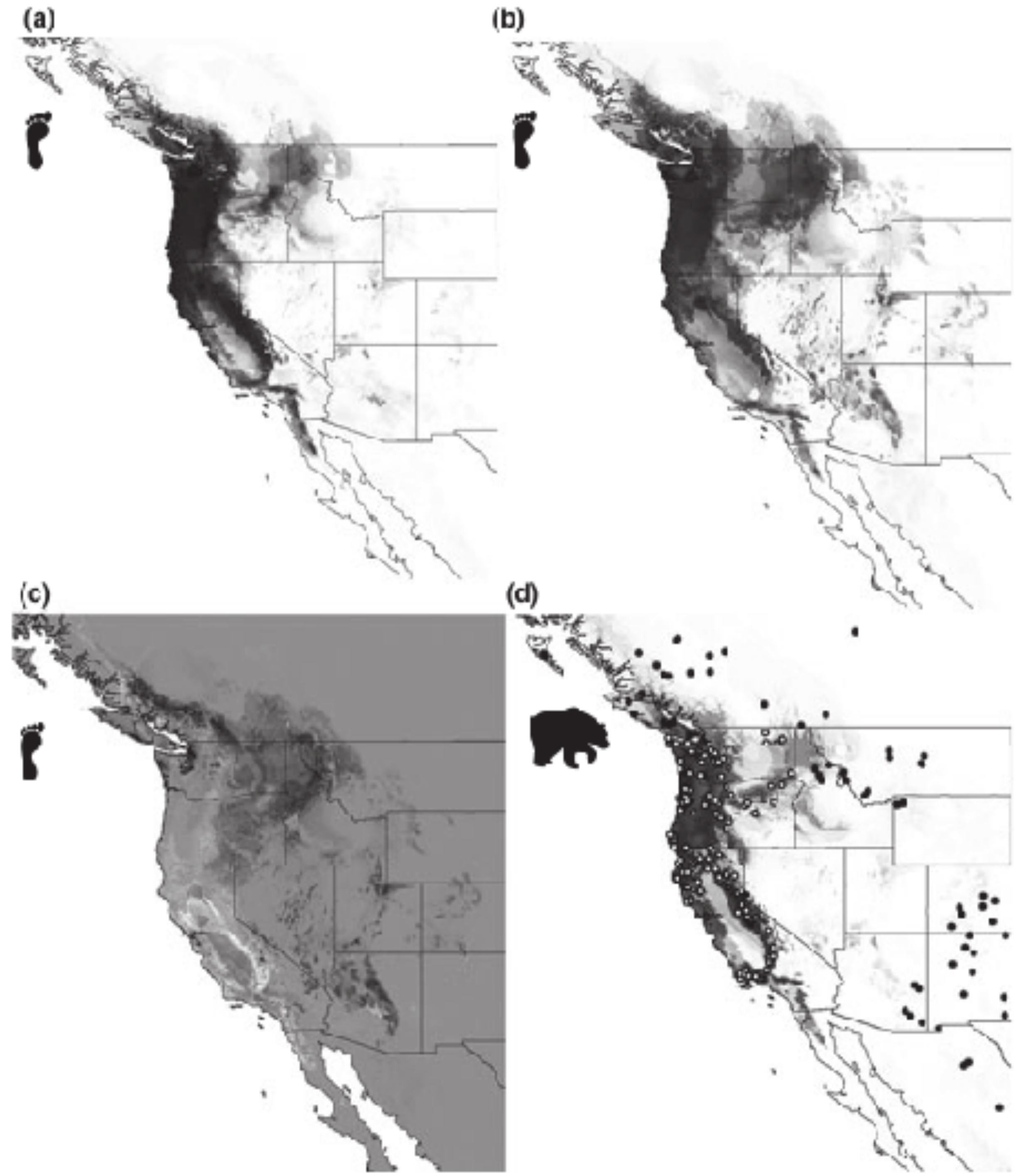


Figure 2 Predicted distributions of Bigfoot constructed from all available encounter data using MAXENT (a) for the present climate and (b) under a possible climate-change scenario involving a doubling of atmospheric CO_2 levels. Results are presented for logistic probabilities of occurrence ranging continuously 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.

PBD: poor quality issues

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

PBD: poor quality issues

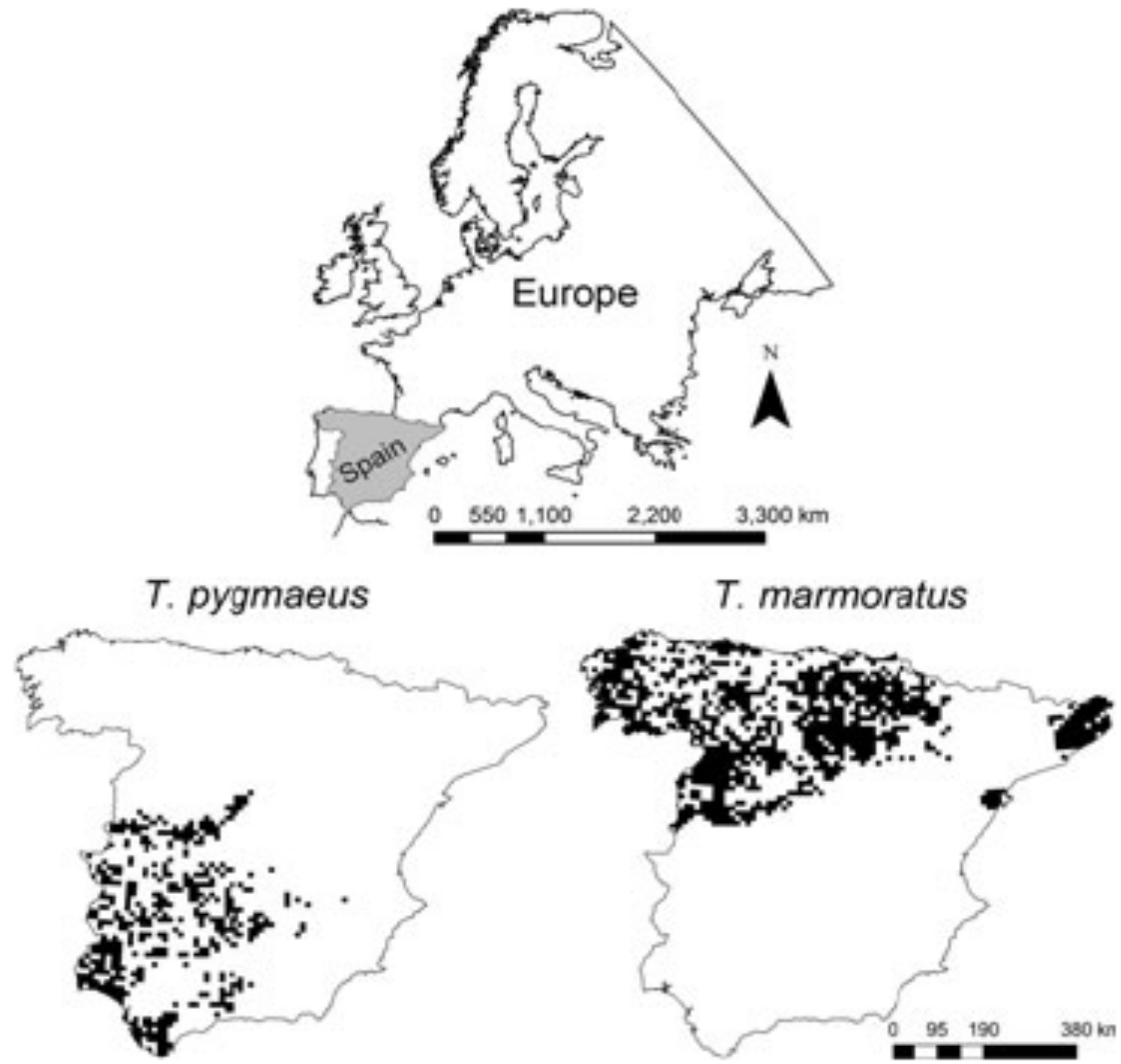


Figure 1 The study area in the European context. Current distributions of *Triturus pygmaeus* and *T. marmoratus* represented in 10 km × 10 km UTM cells within the study area of mainland Spain (data from Pleguezuelos *et al.*, 2004).

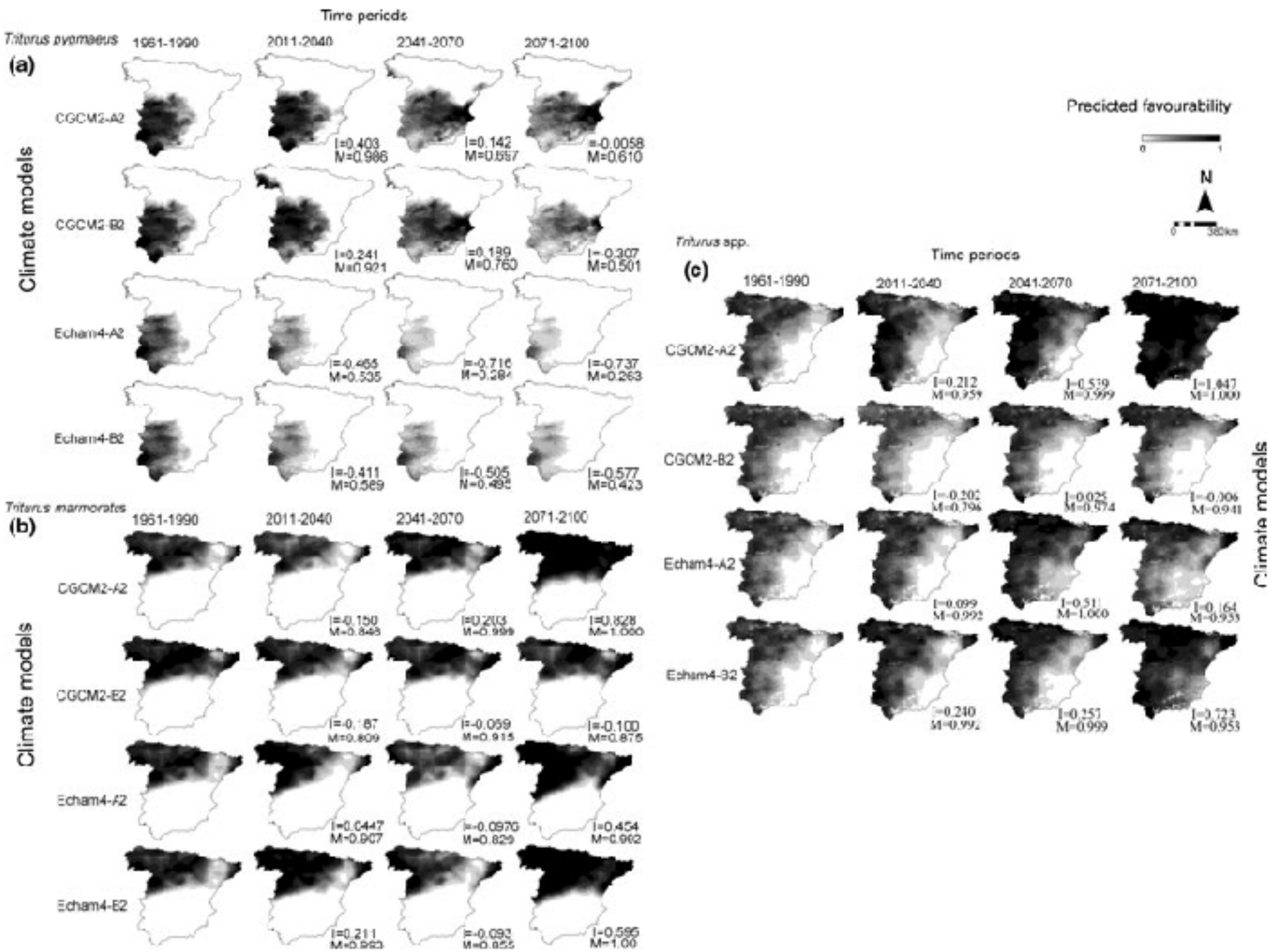


Figure 2 Favourability in mainland Spain predicted for (a) *Triturus pygmaeus* and (b) *T. marmoratus*, as separate species, and (c) for both *Triturus* taxa as belonging to the same species, respectively, according to each climate model and for each time period – ranging from 0 (low favourability) to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

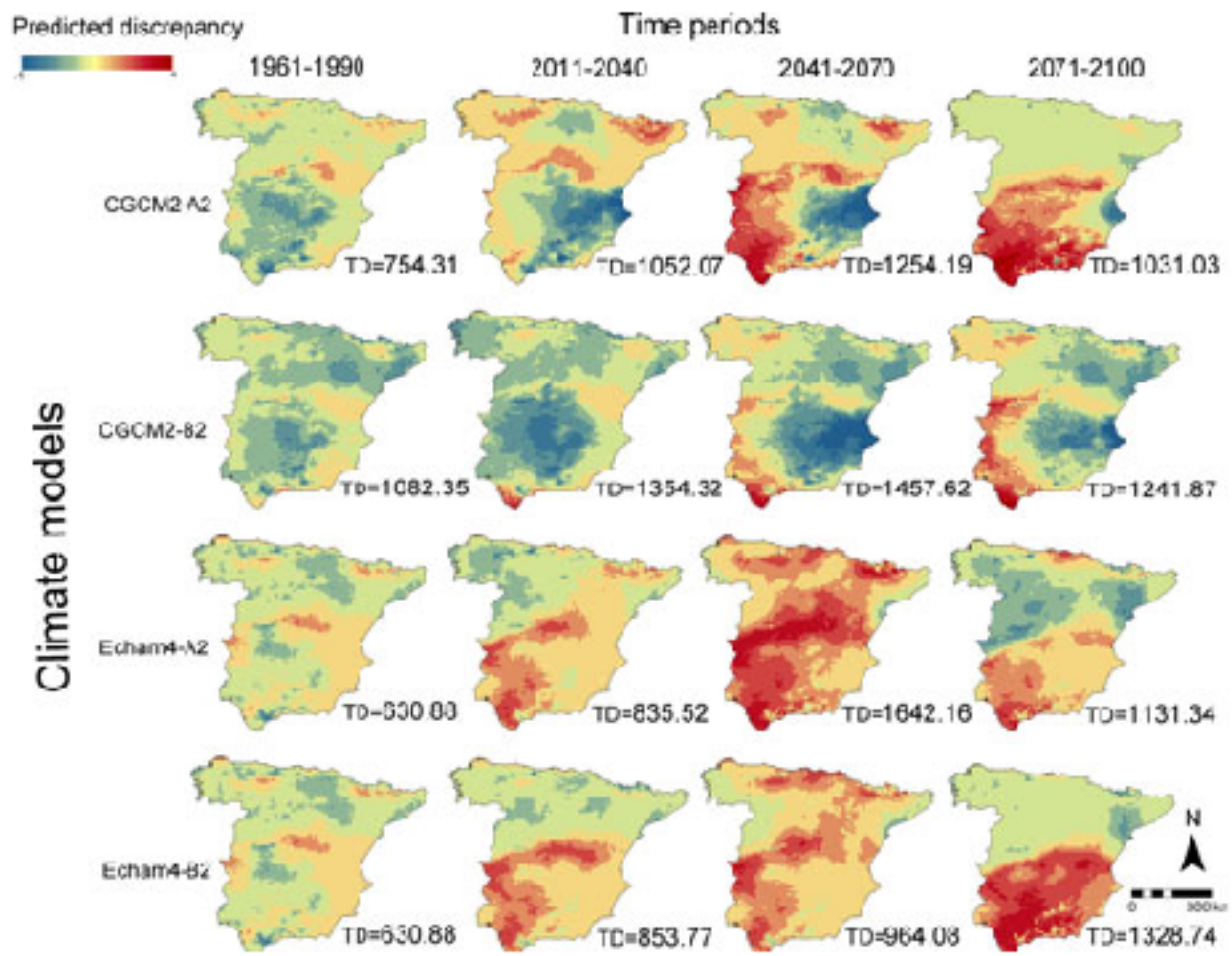


Figure 3 Discrepancy values (difference between favourability predicted for both *Triturus* taxa as belonging to the same species and favourability predicted for the combination of *T. pygmaeus* and *T. marmoratus* as separate species) in mainland Spain according to each climate model and for each time 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, 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.

PBD: poor quality issues

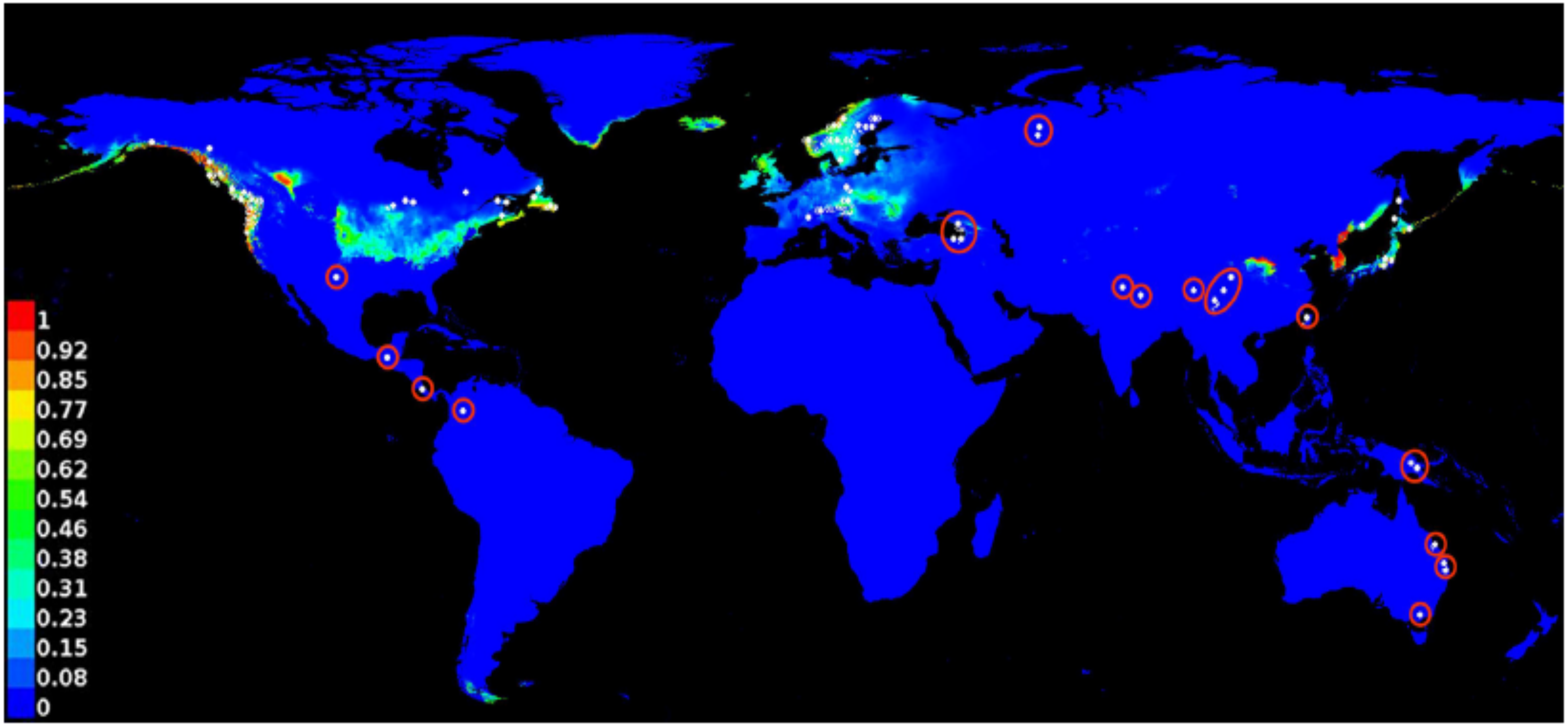


Fig 2. Best fitting MexEnt model for *Usnea longissima* based on 1477 sequenced samples corresponding to 160 localities from Rolstad et al. (2013), indicated by shaded areas ranging from pale blue-green to red. Bright blue areas indicate range of predicted absence. Map is overlayed by occurrence records 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.

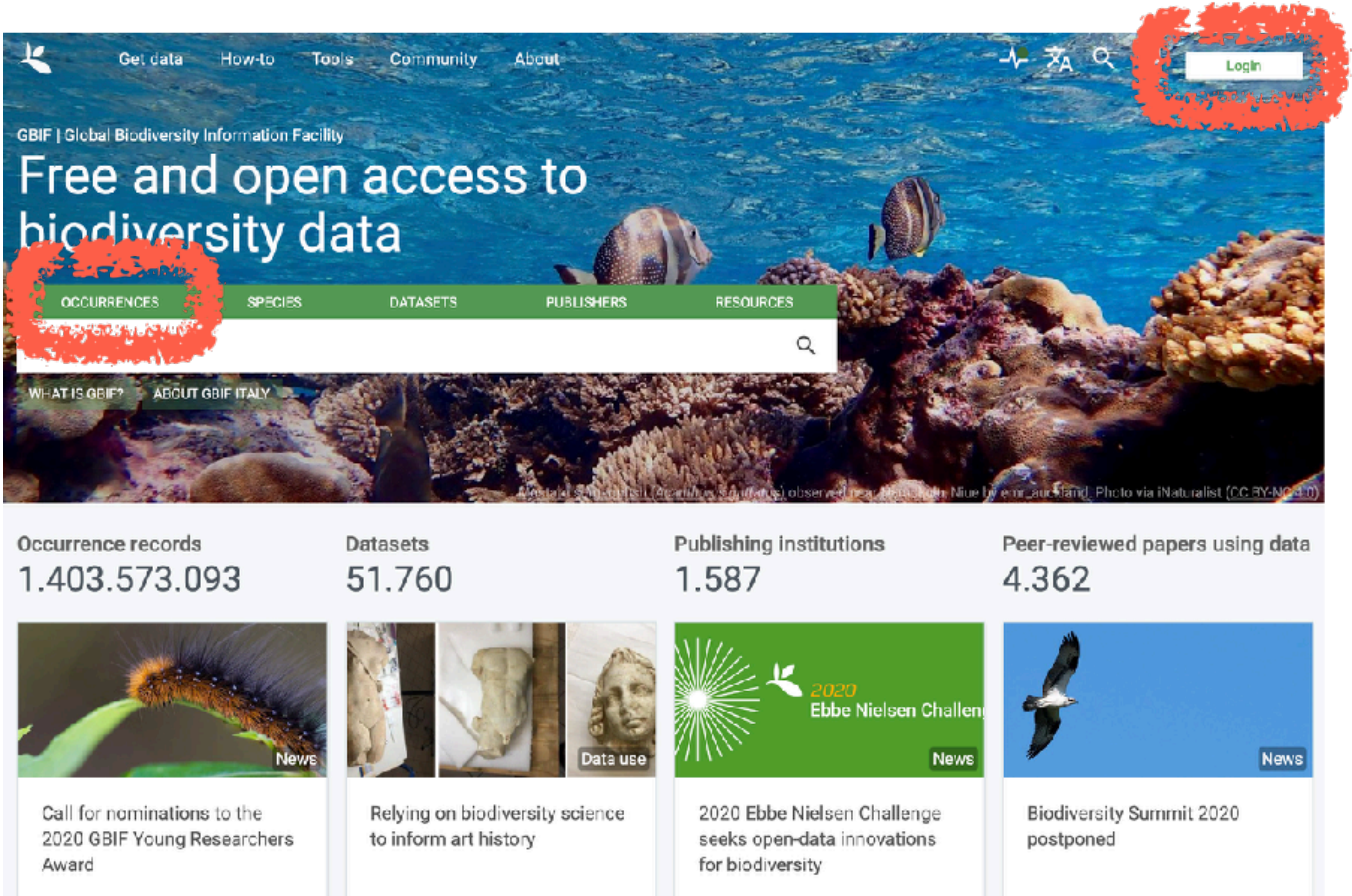
In any case, however, the GBIF - as well as other online data aggregators - remains a fundamental resource for modern research.

Dataset can be obtained from the GBIF in two ways

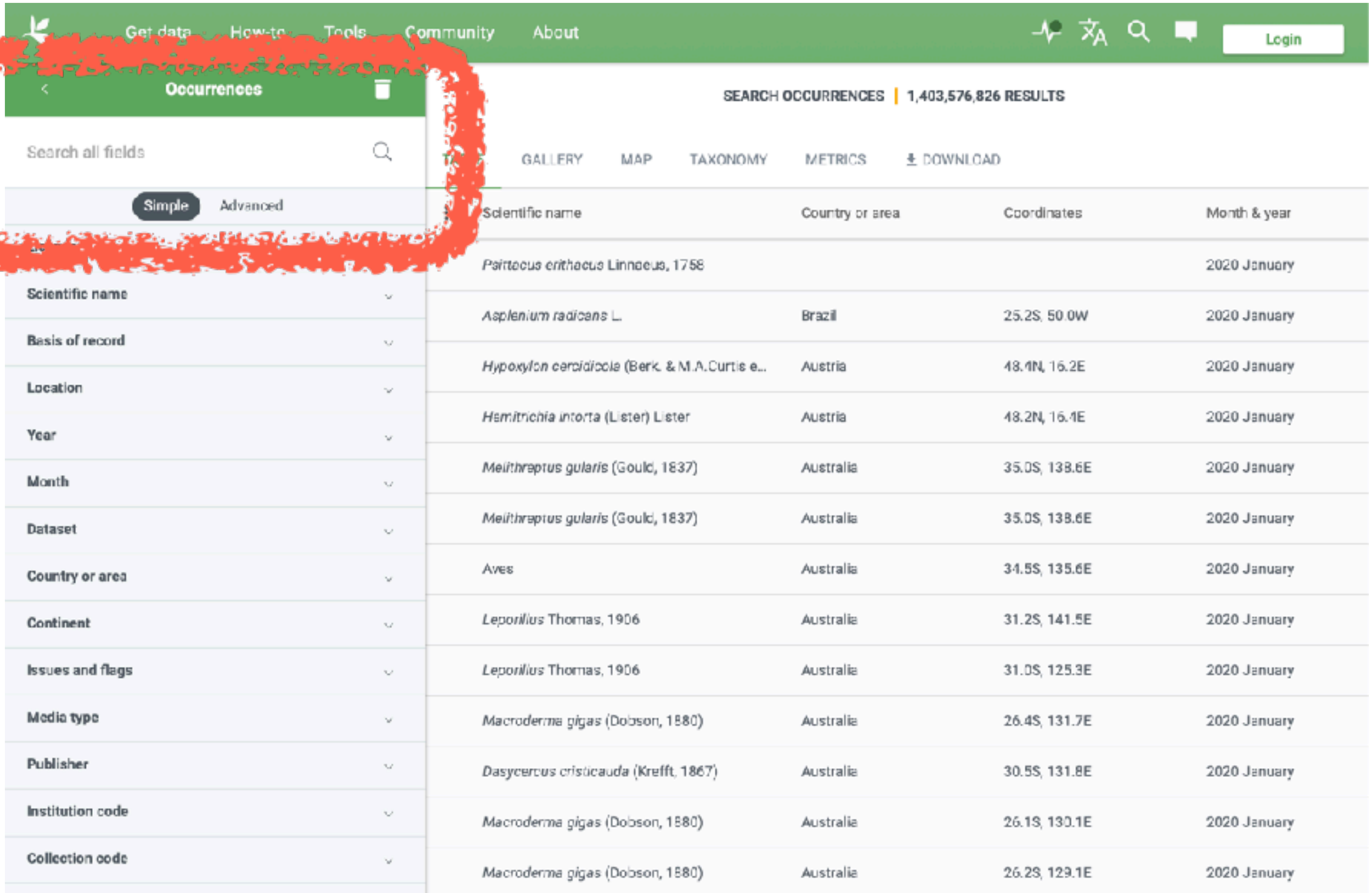
A) by a web browser
B) directly from R

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.

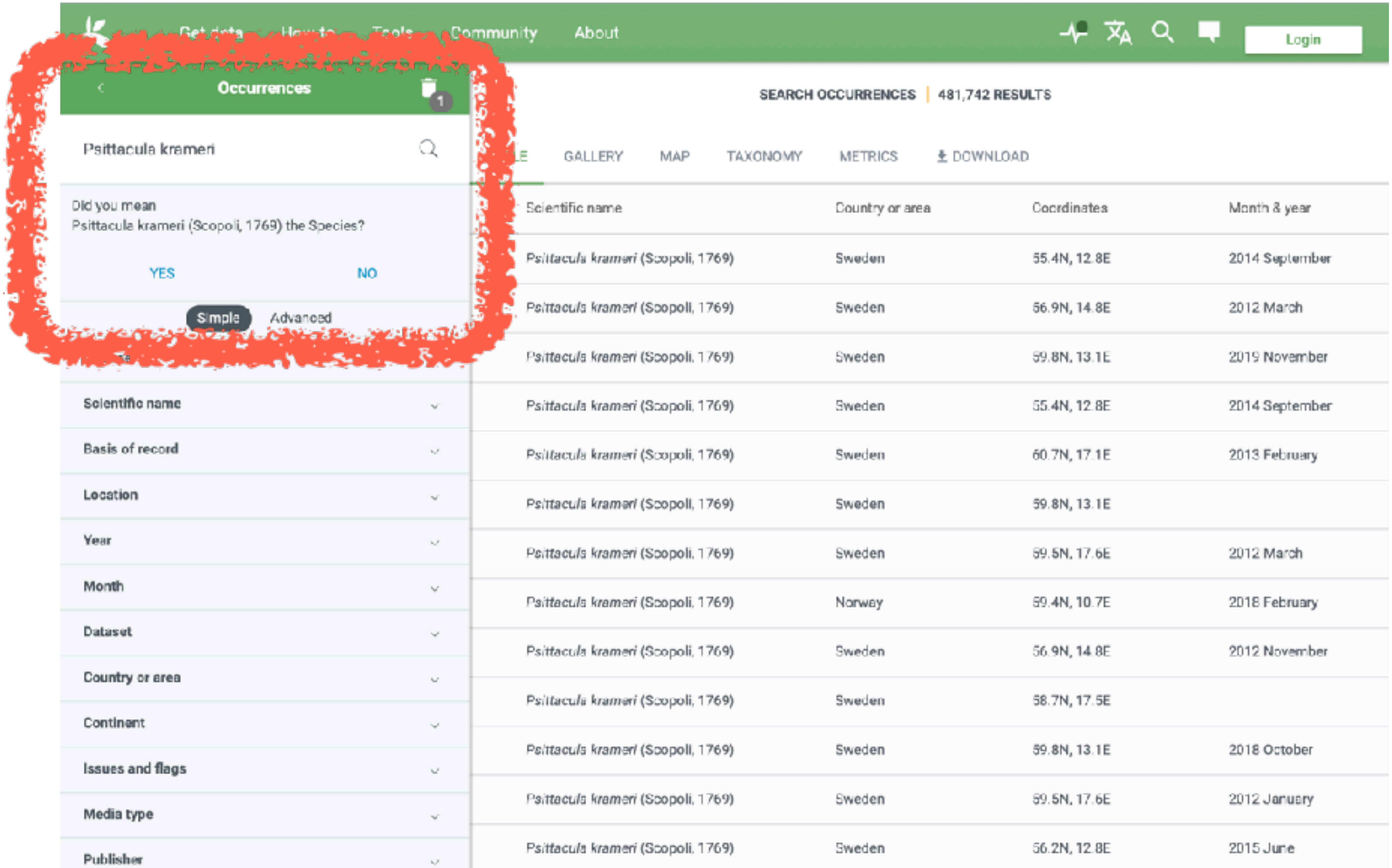
Obtaining PBD



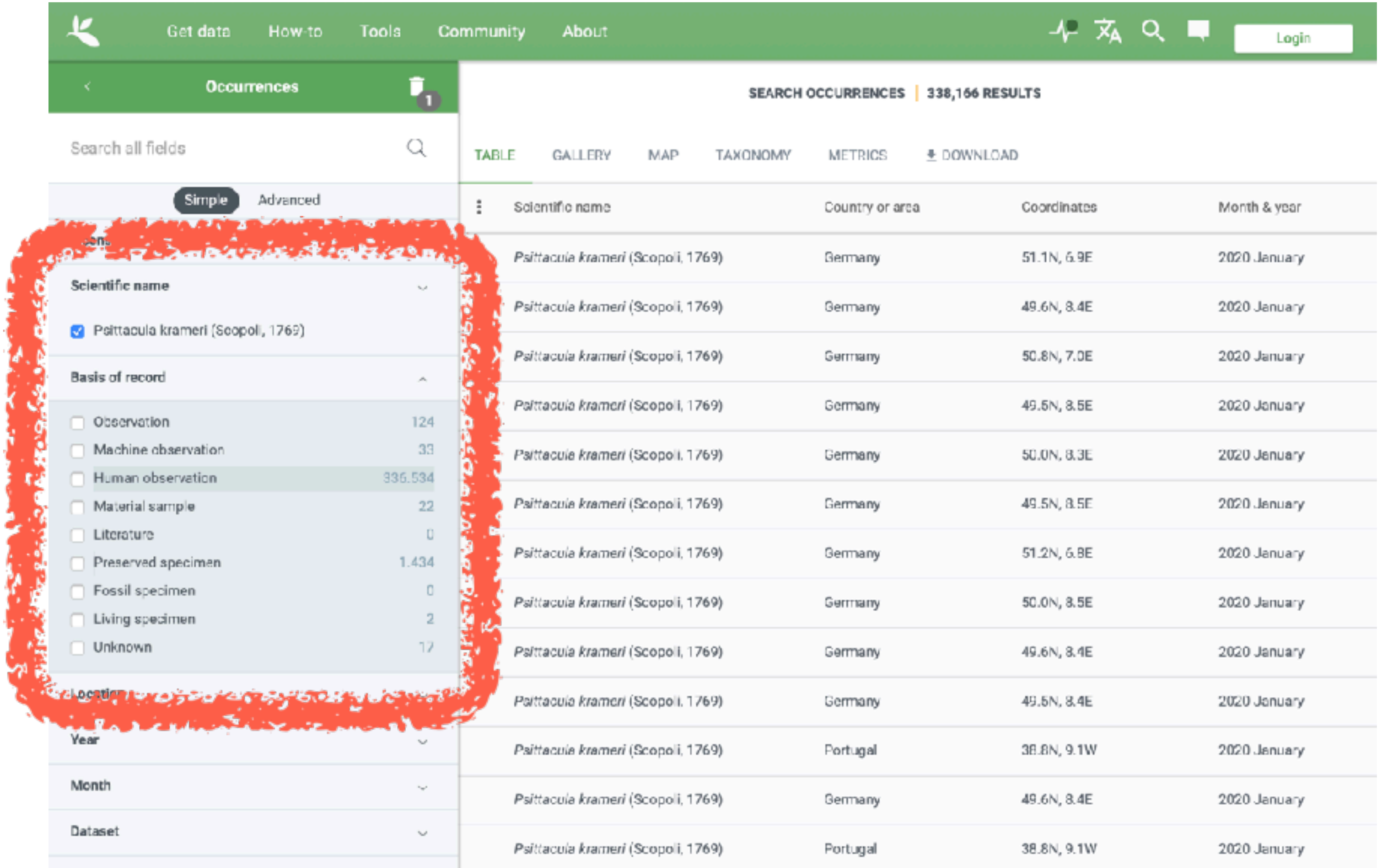
Obtaining PBD



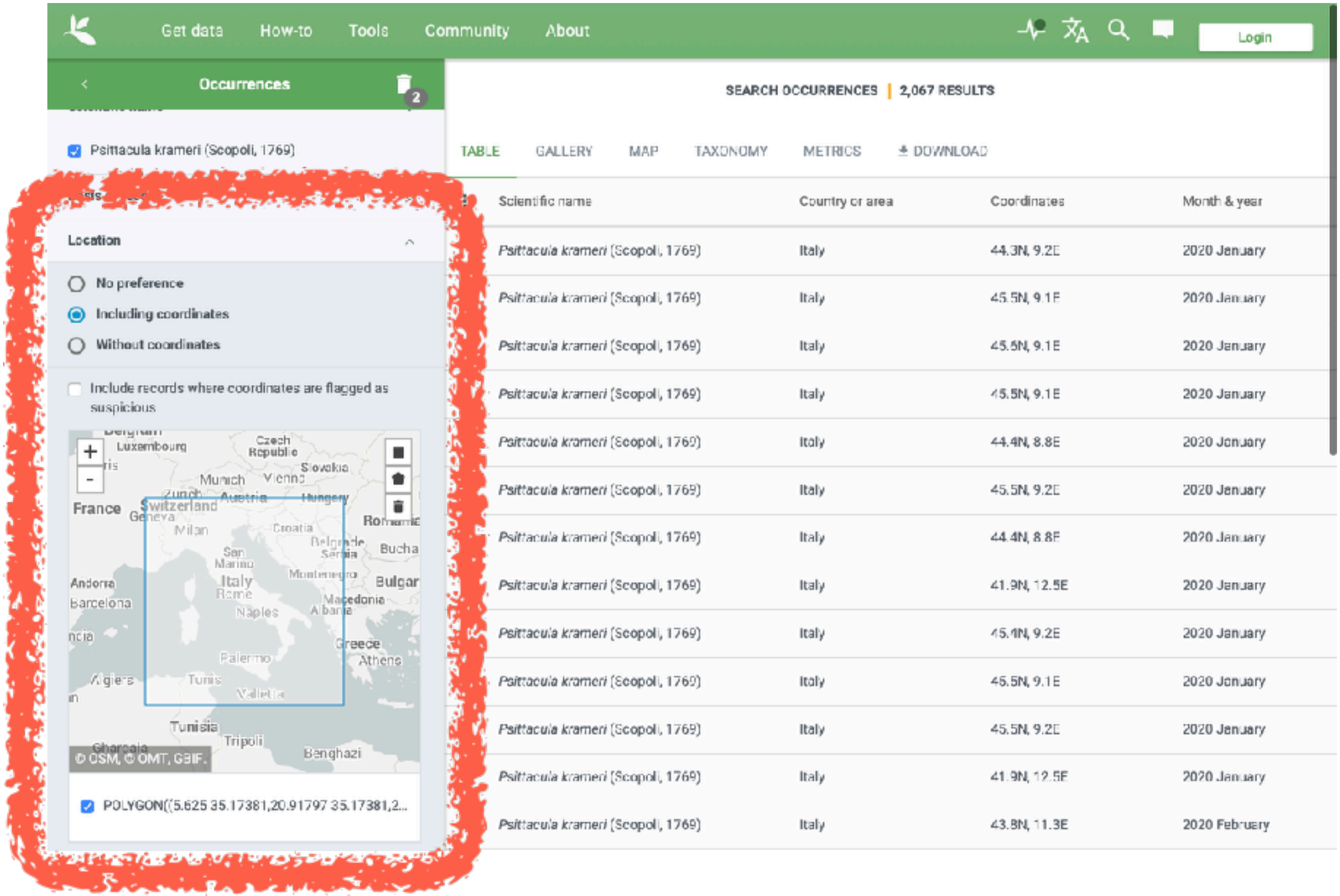
Obtaining PBD



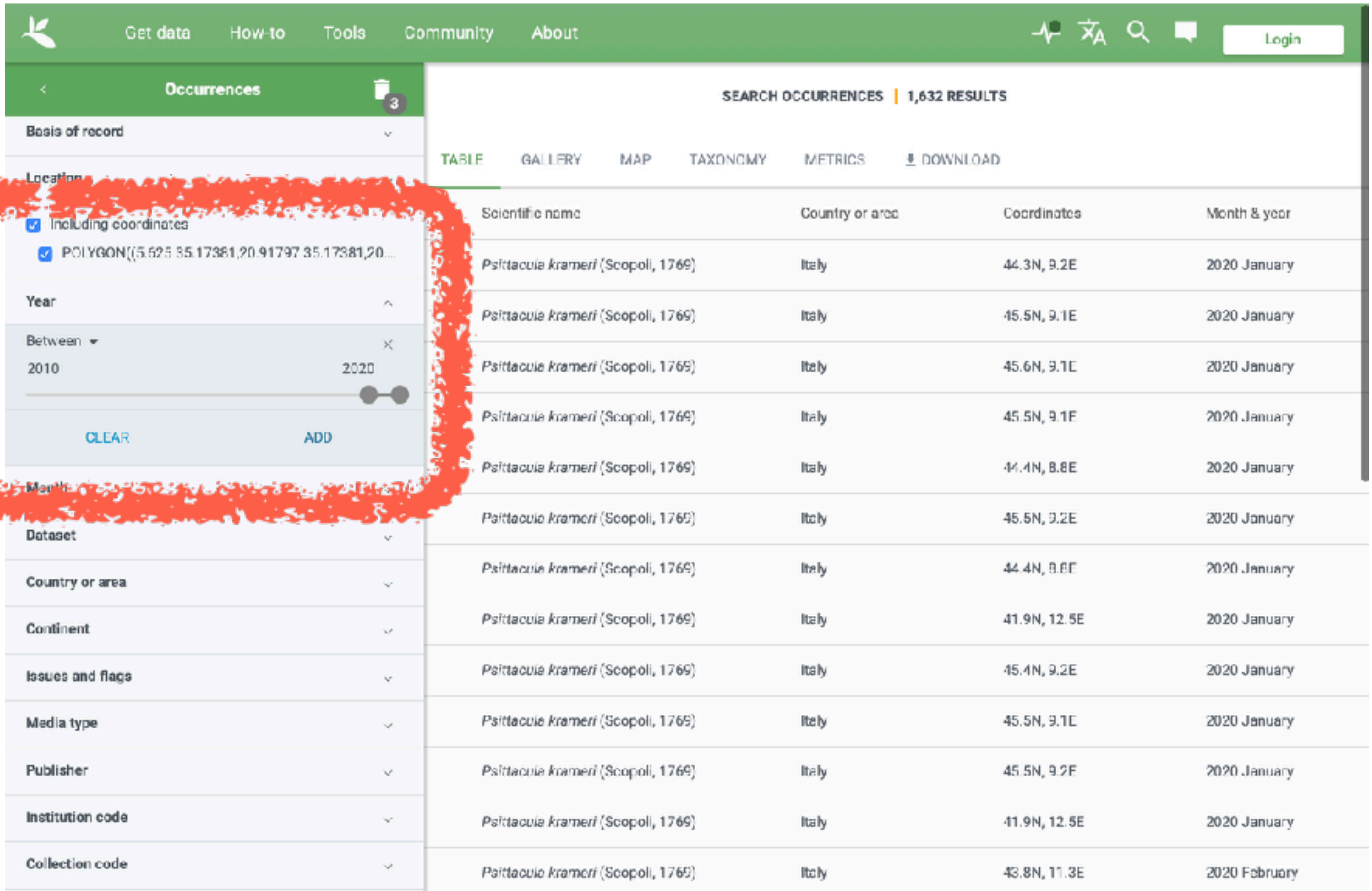
Obtaining PBD



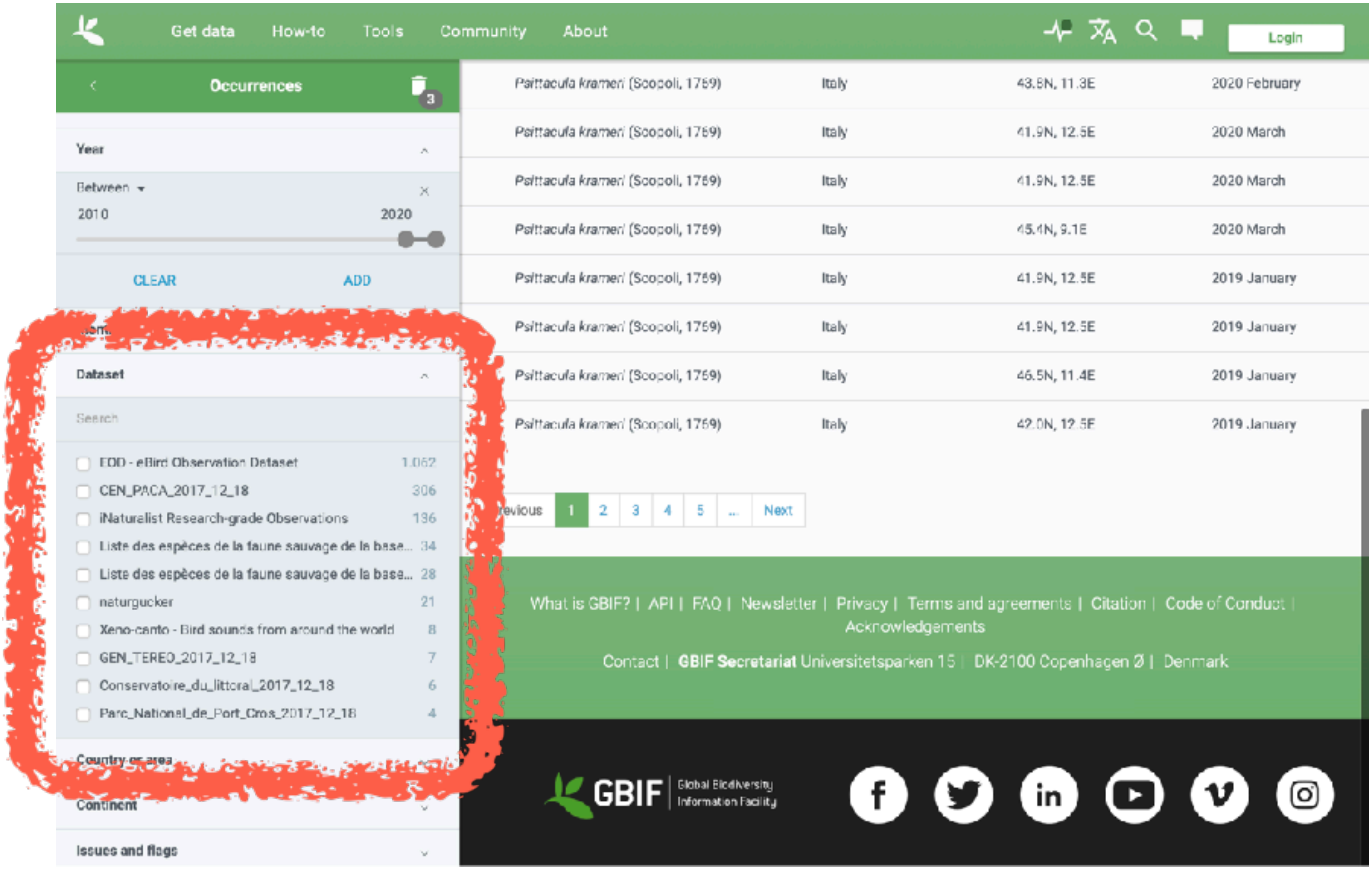
Obtaining PBD

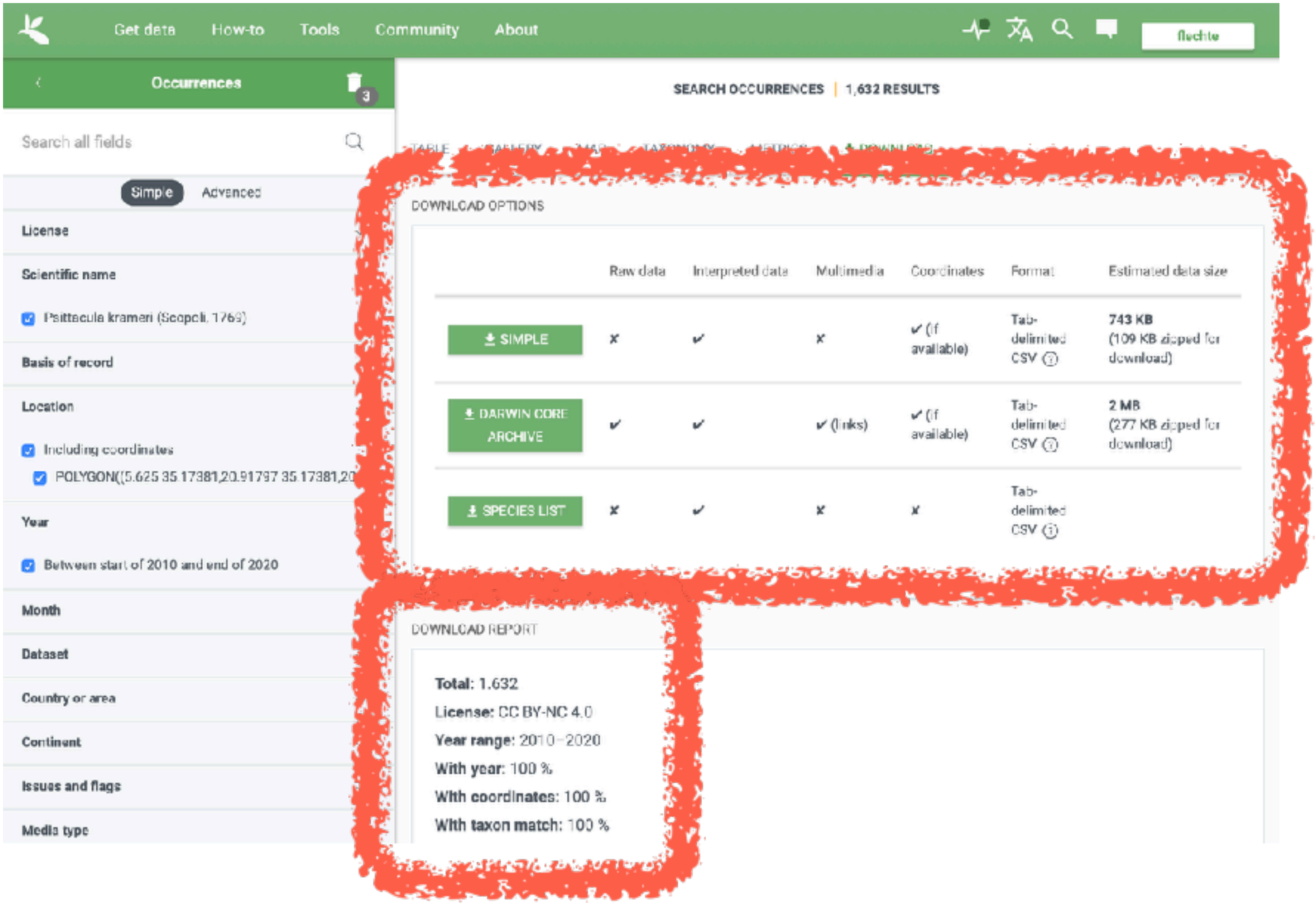


Obtaining PBD

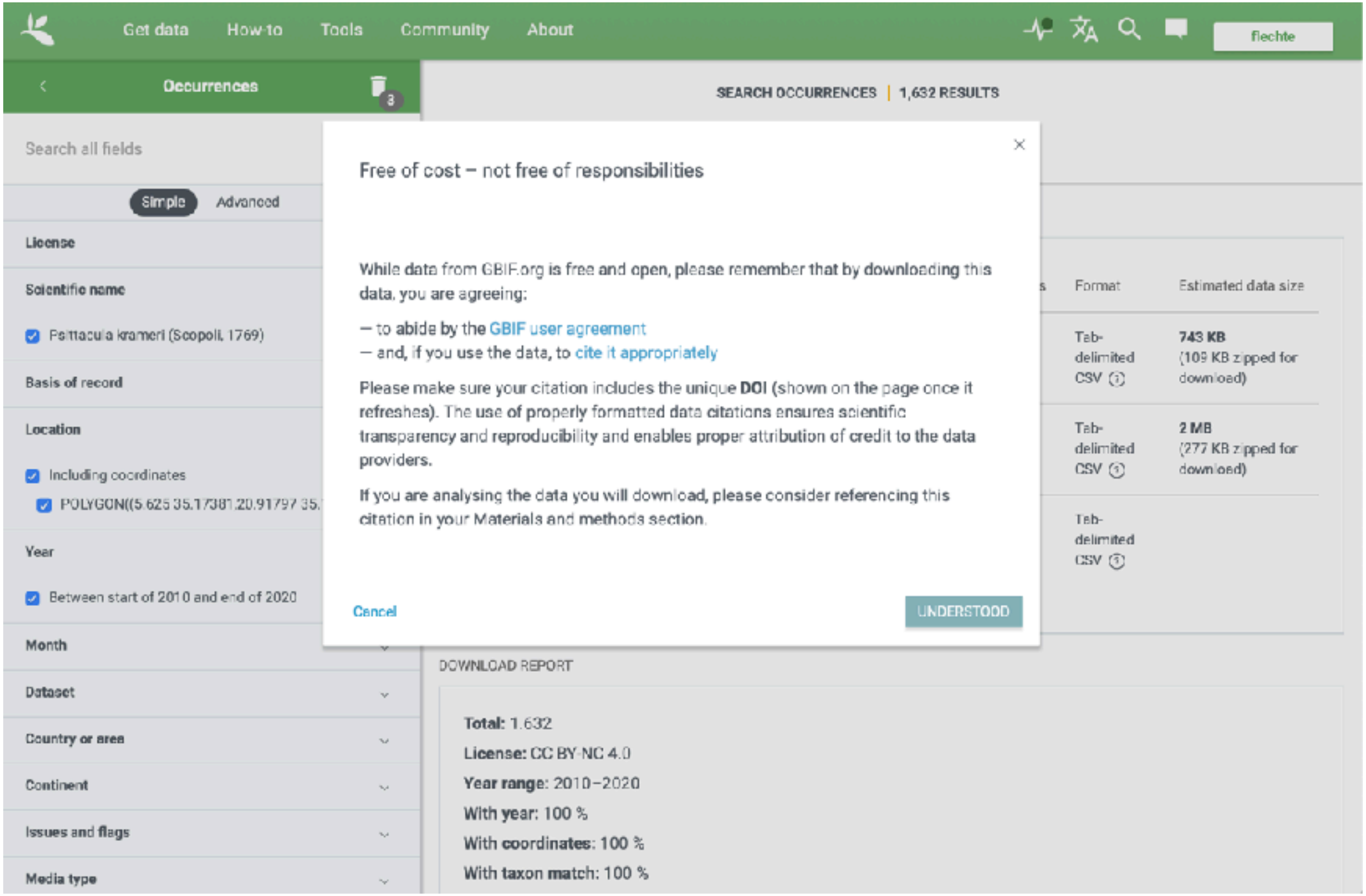


Obtaining PBD

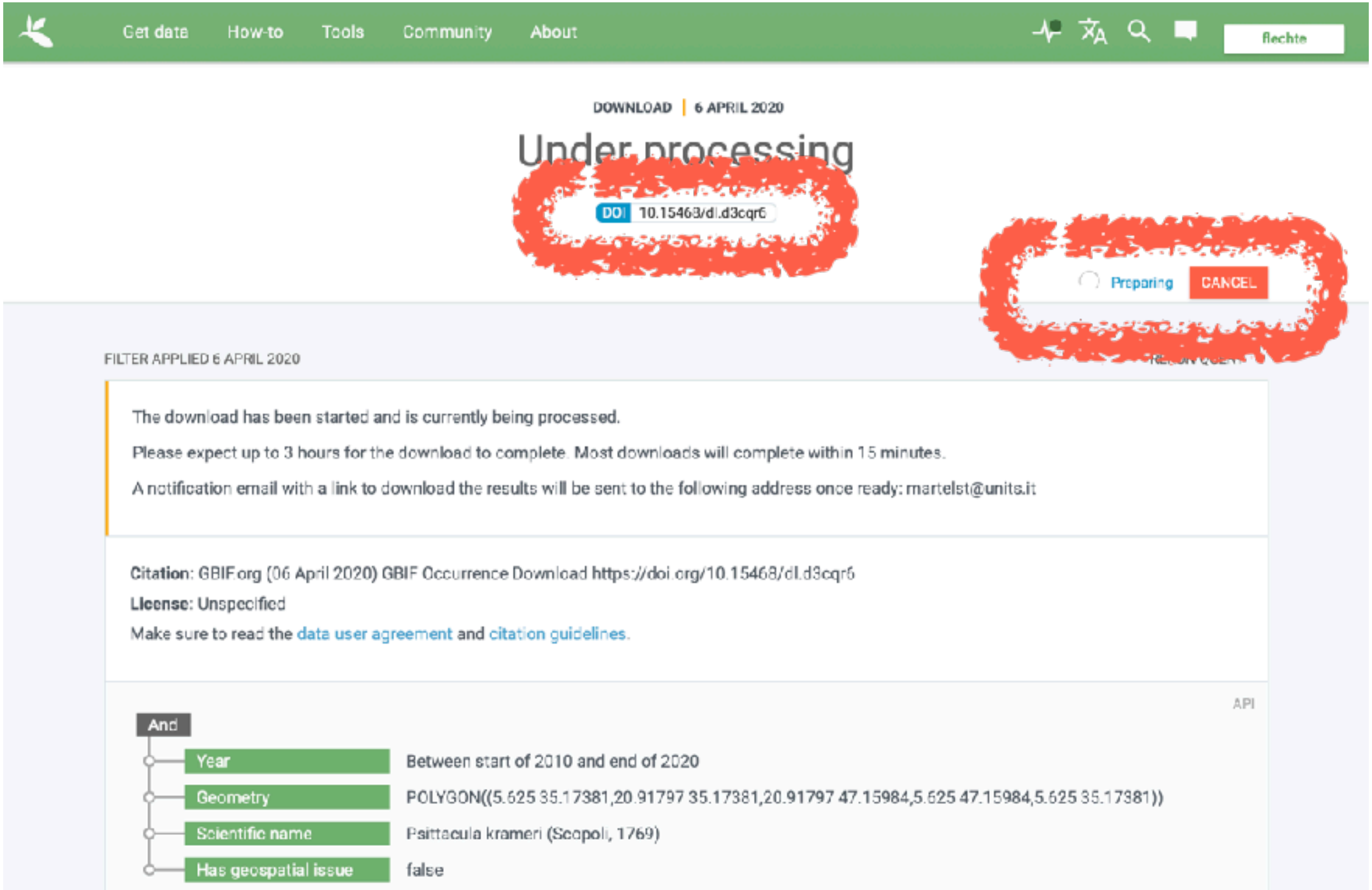



Get data
How-to
Tools
Community
About
Occurrences
SEARCH OCCURRENCES | 1,632 RESULTS
Search all fields
Simple
Advanced
License
Scientific name
Psittacula krameri (Scopoli, 1769)
Basis of record
Location
Including coordinates
POLYGON((5.625 35.17381,20.91797 35.17381,20
Year
Between start of 2010 and end of 2020
Month
Dataset
Country or area
Continent
Issues and flags
Media type
DOWNLOAD OPTIONS
Raw data
Interpreted data
Multimedia
Coordinates
Format
Estimated data size
SIMPLE
✔ (if available)
Tab-delimited CSV
743 KB (109 KB zipped for download)
DARWIN CORE ARCHIVE
✔ (links)
✔ (if available)
Tab-delimited CSV
2 MB (277 KB zipped for download)
SPECIES LIST
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 %

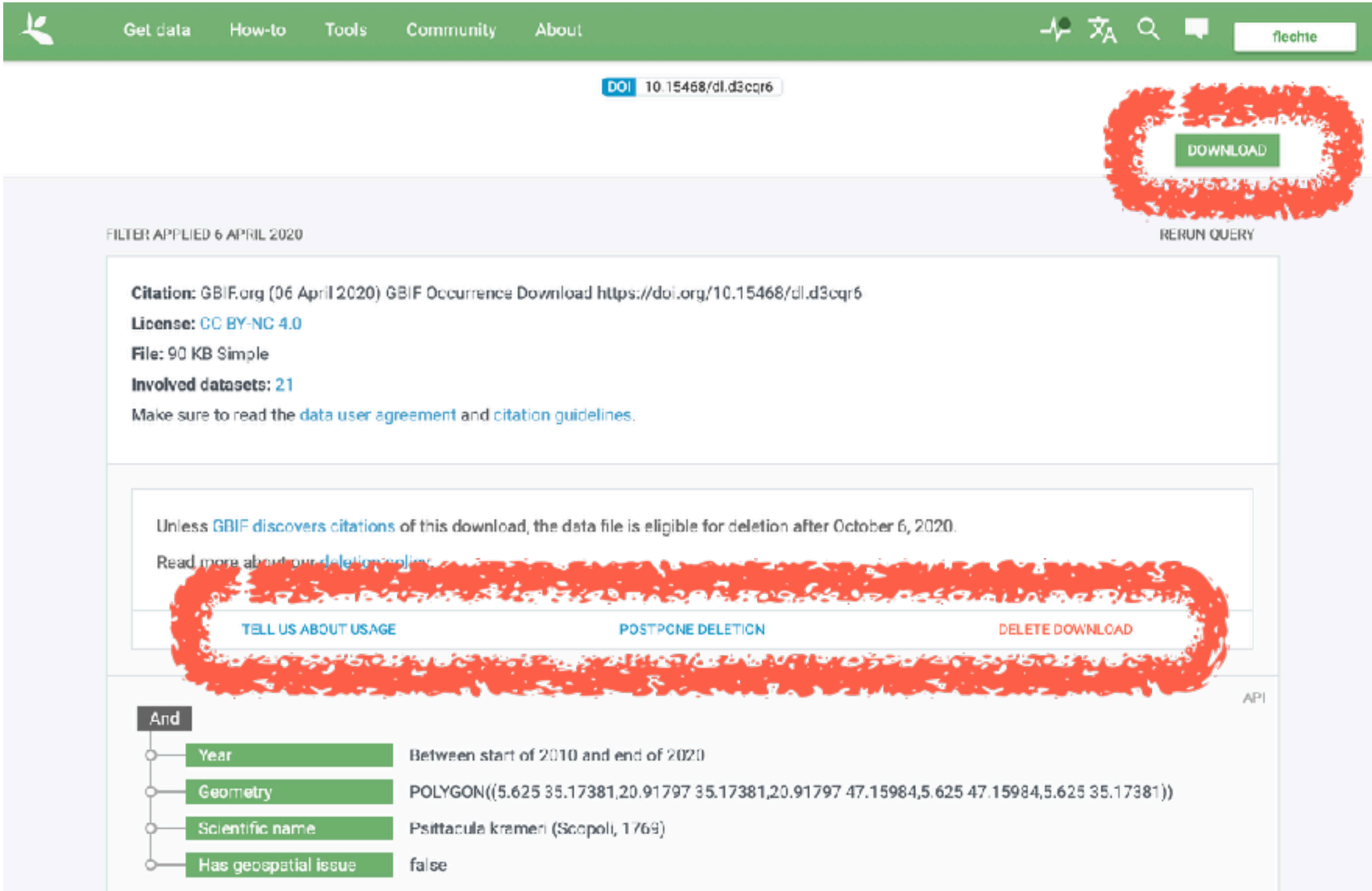
Obtaining PBD



Obtaining PBD



Obtaining PBD



downloads@gbif.org 09:24 D

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/request/0035948-200221144449610.zip

Link to the file

When using this dataset please use the following citation:
GBIF.org (6 April 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.d3cqr6

Citation

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

Recap of query parameters

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.

File retention

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

Obtaining PBD

gbifID	datasetKey	occurrenceID	kingdom	phylum	class	order	family	genus	species
2580194357	50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/39944292	Animalia	Chordata	Aves	Psittaciformes	Psittacidae	Psittacula	Psittacula krameri
2580125218	50c9509d-22c7-4a22-a47d-8c48425ef4a7	https://www.inaturalist.org/observations/39969968	Animalia	Chordata	Aves	Psittaciformes	Psittacidae	Psittacula	Psittacula krameri
2580									
2579									
2578									
2578									
2576									
2574									
2574									
2573									
2573									
2573									
2563									
2563									
2563									
2563									
2563									
2557									
2557									
2557									
2557									
2550									
2549									
2543									
2542									
2540									
2540									

infraspecificEpithet	taxonRank	scientificName	verbatimScientificName	verbatimScientificNameAuthorship	countryCode	locality	stateProvince
	SPECIES	Psittacula krameri (Scopoli, 1769)	Psittacula krameri		IT		Lombardia
	SPECIES	Psittacula krameri (Scopoli, 1769)	Psittacula krameri		IT		Lazio

occurrenceStatus	individualCount	publishingOrgKey	decimalLatitude	decimalLongitude	coordinateUncertaintyInMeters	coordinatePrecision	elevation	elevationAccuracy
		28eb1a3f-1c15-4a95-931a-4af90ecb574d	45.426071	9.139852	1379.0			
		28eb1a3f-1c15-4a95-931a-4af90ecb574d	41.852145	12.50296	17.0			

depth	depthAccuracy	eventDate	day	month	year	taxonKey	speciesKey	basisOfRecord	institutionCode	collectionCode	catalogNumber	recordNumber
		2020-03-13T16:55:00Z	13	3	2020	2479226	2479226	HUMAN_OBSERVATION	iNaturalist	Observations	39944292	
		2020-03-14T16:13:03Z	14	3	2020	2479226	2479226	HUMAN_OBSERVATION	iNaturalist	Observations	39969968	

identifiedBy	dateIdentified	license	rightsHolder	recordedBy	typeStatus	establishmentMeans	lastInterpreted
	2020-03-13T23:14:44Z	CC_BY_4_0	Alexandros Quartarone	Alexandros Quartarone			2020-03-30T20:24:53.393Z
	2020-03-14T16:18:20Z	CC_BY_NC_4_0	marco caporioni	marco caporioni			2020-03-30T20:24:11.745Z

typeStatus	establishmentMeans	lastInterpreted	mediaType	issue
		2020-03-30T20:24:53.393Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:11.745Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:32.367Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:19.066Z		COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:58.480Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:23.463Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:12.286Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:03.881Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:04.275Z		COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:52.032Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:18.370Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:04.612Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:23:56.852Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:31.788Z		COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:09.968Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:37.158Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:23:57.586Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:02.705Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:14.931Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:12.862Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:48.424Z	StillImage	GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:37.663Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:03.459Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:52.456Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:25:30.342Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:23:58.470Z	StillImage	COORDINATE_ROUNDED;GEODETIC_DATUM_ASSUMED_WGS84
		2020-03-30T20:24:40.112Z	StillImage	COORDINATE_ROUNDED;GEODETIC_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