

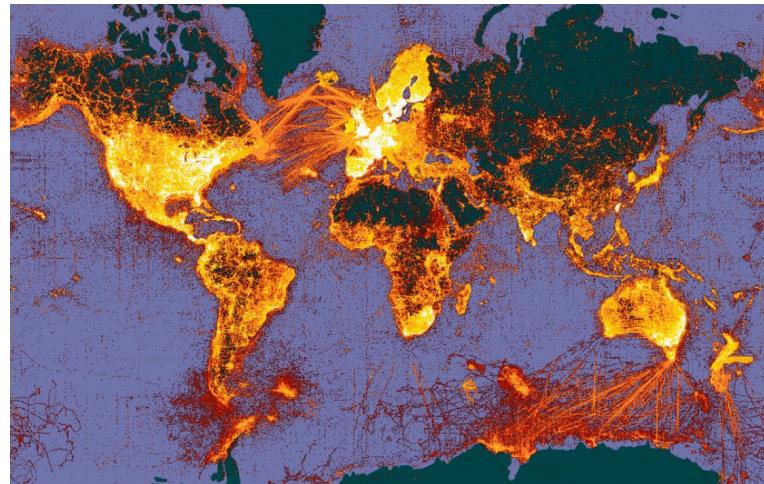


Биоәртүрлілікті зерттеудегі цифрлық технологиялар
Digital technologies in biodiversity research
Цифровые технологии в исследовании биоразнообразия



ЛЕКЦИЯ 6

Портал GBIF -
крупнейший мировой
репозиторий открытых
данных о
биоразнообразии



Слайды CC BY:

Dag Endresen, GBIF Norway

Наталья Иванова,

Максим Шашков

Репозитории

Get data Share Tools Inside GBIF

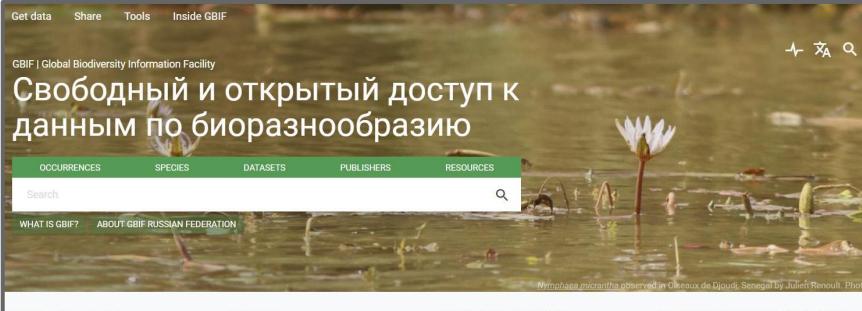
GBIF | Global Biodiversity Information Facility

Свободный и открытый доступ к данным по биоразнообразию

OCCURRENCES SPECIES DATASETS PUBLISHERS RESOURCES

Search 

WHAT IS GBIF? ABOUT GBIF RUSSIAN FEDERATION



Nymphaea micrantha observed in Oiseaux de Djoudi, Senegal by Julien Renoult. Photo

Occurrence records 1 339 174 962 Datasets 46 267 Publishing institutions 1 464 Peer-reviewed papers using data 3 896



Five projects receive funding from 2019 Capacity Enhancement Support Programme
14 August 2019



Predicting the future of biodiversity using Essential Biodiversity Variables
12 September 2019



Belarus extends GBIF's European membership map eastward
16 July 2019



Programme seeks Biodiversity Open Data Ambassadors to expand best practices
10 July 2019

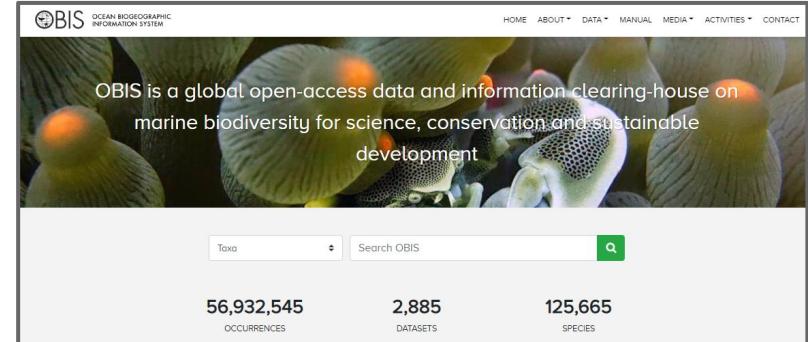
OBIS OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

HOME ABOUT DATA MANUAL MEDIA ACTIVITIES CONTACT

OBIS is a global open-access data and information clearing-house on marine biodiversity for science, conservation and sustainable development

Toxu  Search OBIS 

56,932,545 OCCURRENCES 2,885 DATASETS 125,665 SPECIES



BOLD SYSTEMS

DATABASES IDENTIFICATION TAXONOMY WORKBENCH RESOURCES LOGIN 

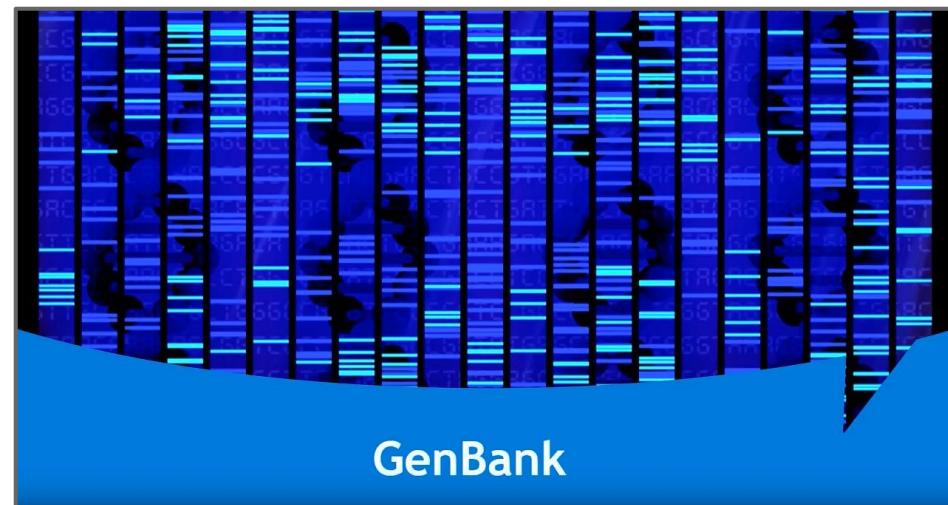
BARCODE OF LIFE DATA SYSTEM v4

Advancing biodiversity science through DNA-based species identification.

EXPLORE THE DATA

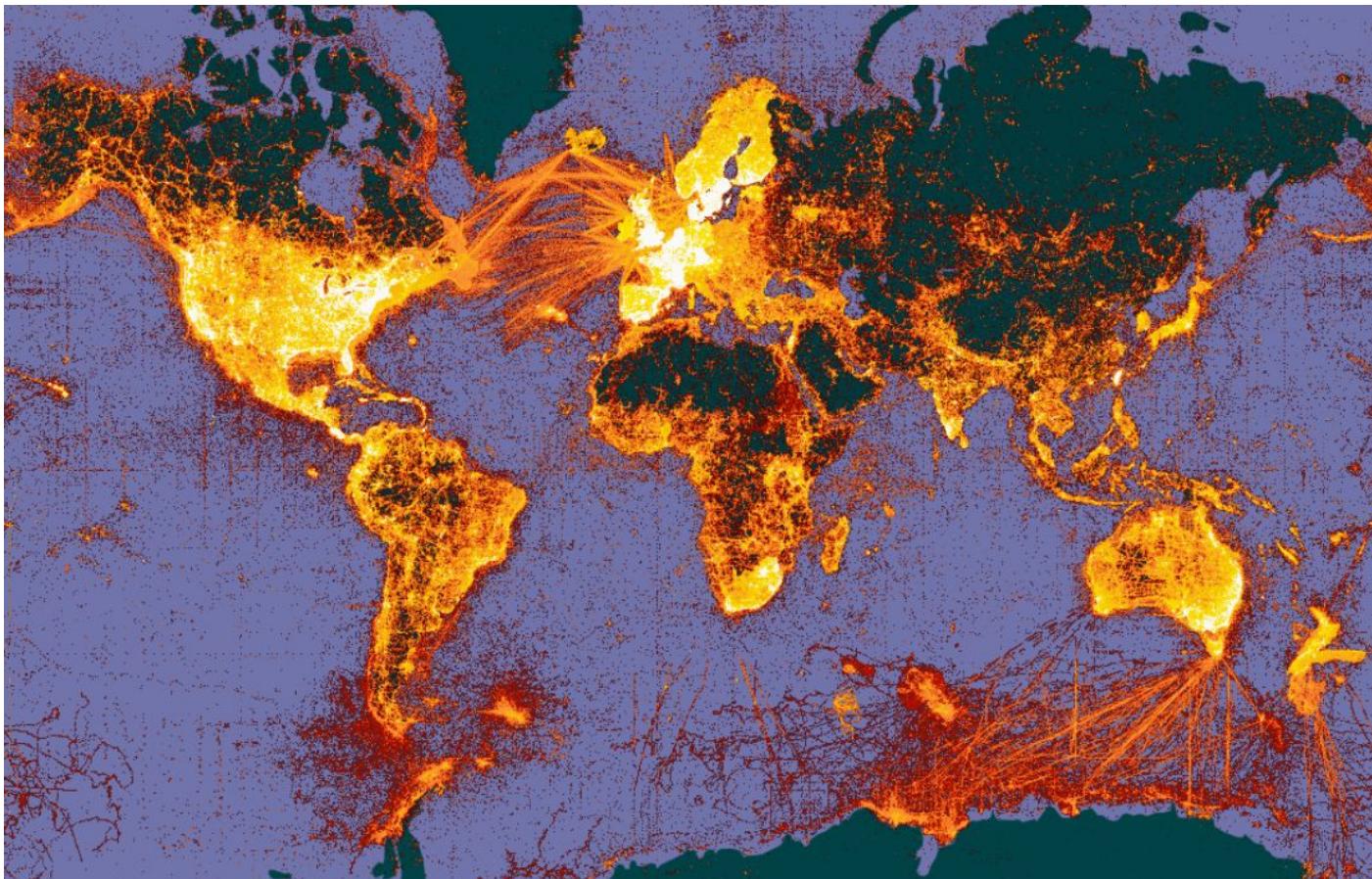
DESIGNED TO SUPPORT THE GENERATION & APPLICATION OF DNA BARCODE DATA

BOLD is a cloud-based data storage and analysis platform developed at the Centre for Biodiversity Genomics in Canada. It consists of four main modules, a data portal, an educational portal, a registry of BINs (putative species), and a data collection and analysis workbench.



GenBank

GBIF - крупнейший агрегатор данных о биоразнообразии



1 339 174 017
находок

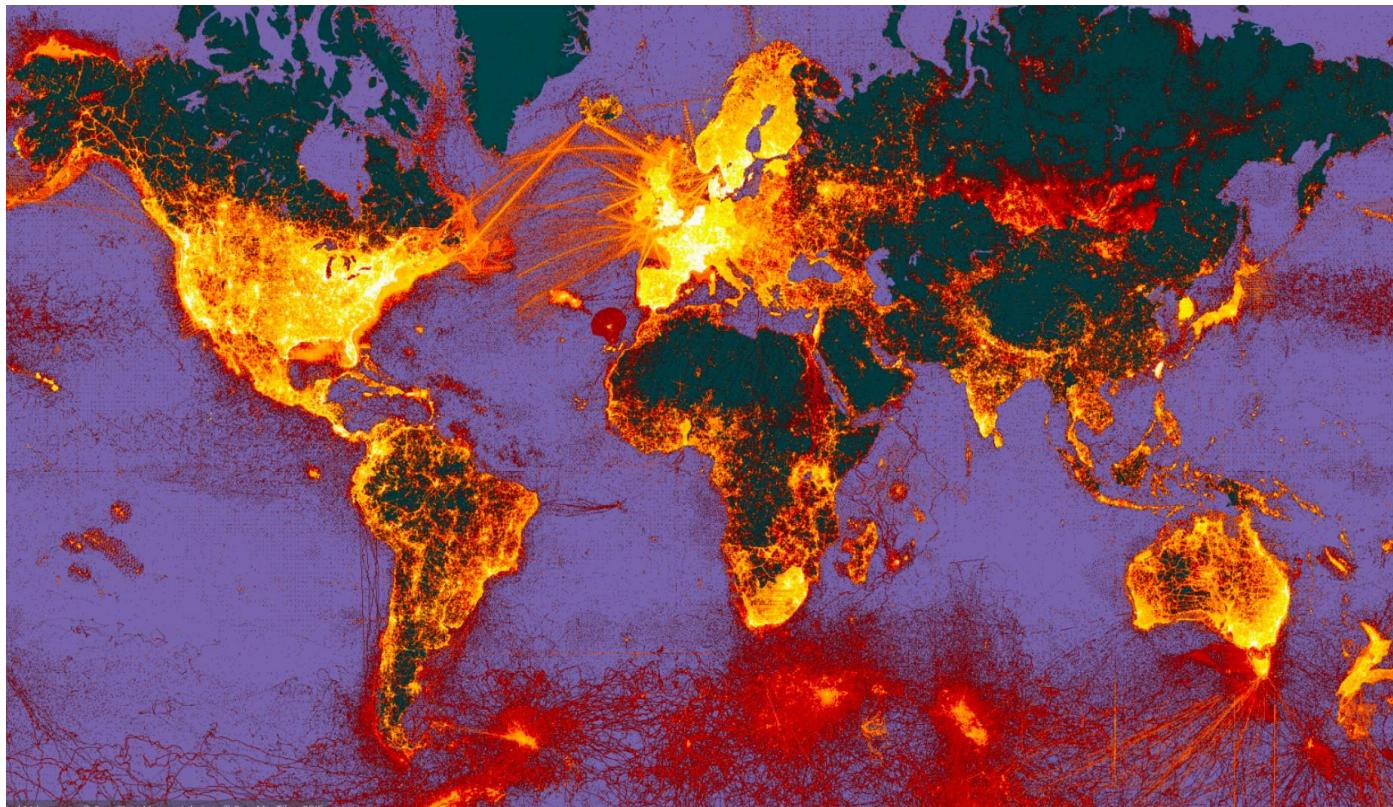
46 267
наборов данных

1464
организаций

3896 статей с
использованием
данных GBIF

15 сентября 2019 г.

GBIF - крупнейший агрегатор данных о биоразнообразии



2 246 485 892
находок

79 054
наборов данных

1926
организаций

7902 статей с
использованием
данных GBIF

1 ноября 2021 г.

Казахстан в GBIF

DATA ABOUT KAZAKHSTAN

337,616

Occurrences

756

Datasets

39

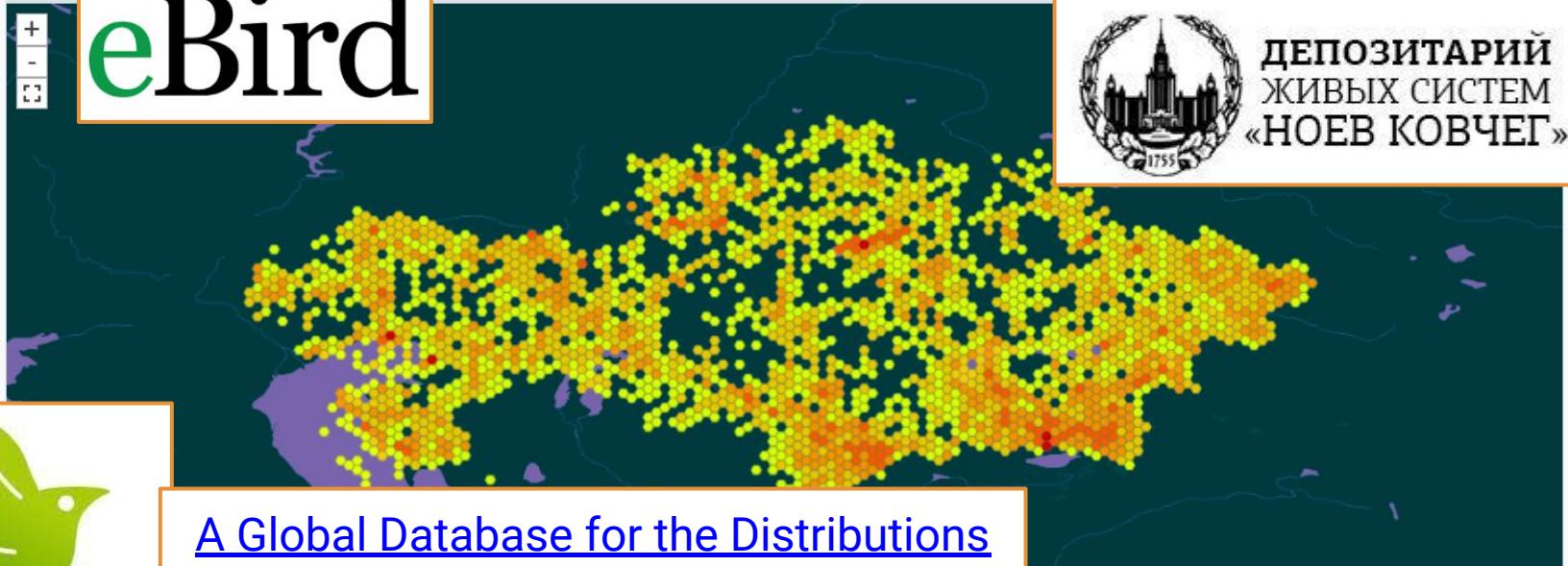
Countries and areas contribute data

252

Publishers



eBird



ДЕПОЗИТАРИЙ
ЖИВЫХ СИСТЕМ
«НОЕВ КОВЧЕГ»



iNaturalist

[A Global Database for the Distributions
of Crop Wild Relatives](#)

2022

[Hymenoptera Specimen
Database of Kyushu University](#)

Казахстан в GBIF

DATA FROM KAZAKHSTAN

94,762

Published occurrences

1

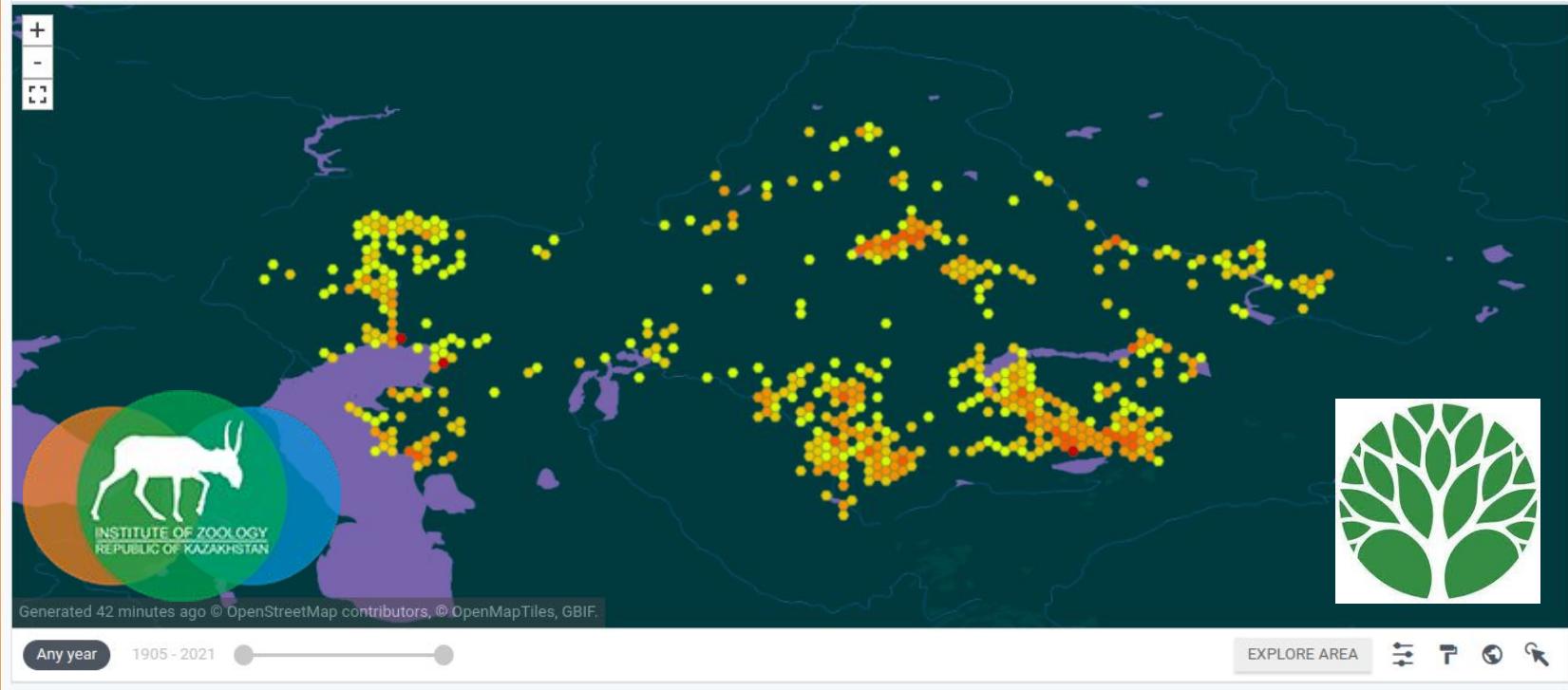
Published datasets

1

Countries and areas covered by data from Kazakhstan

2

Publishers from Kazakhstan



Первый набор данных

OCCURRENCE DATASET | REGISTERED OCTOBER 13, 2022

Distribution of marsh frogs (*Pelophylax ridibundus* complex) in Kazakhstan

Published by [Institute of Zoology of the Republic of Kazakhstan](#)

Dujsebayeva T • Kaptyonkina A • Arifulova I • Ualiyeva D • Akhmedenov K • Ivanov A • Khromov V • Krainyuk V • Sarzhanov F • Tarasovskaya N • Titov S • Timoshenko A • Ermakov O • Malakhov D • Starikov S • Morozov V

DATASET

METRICS

ACTIVITY

 DOWNLOAD

 HOME PAGE

110 OCCURRENCES

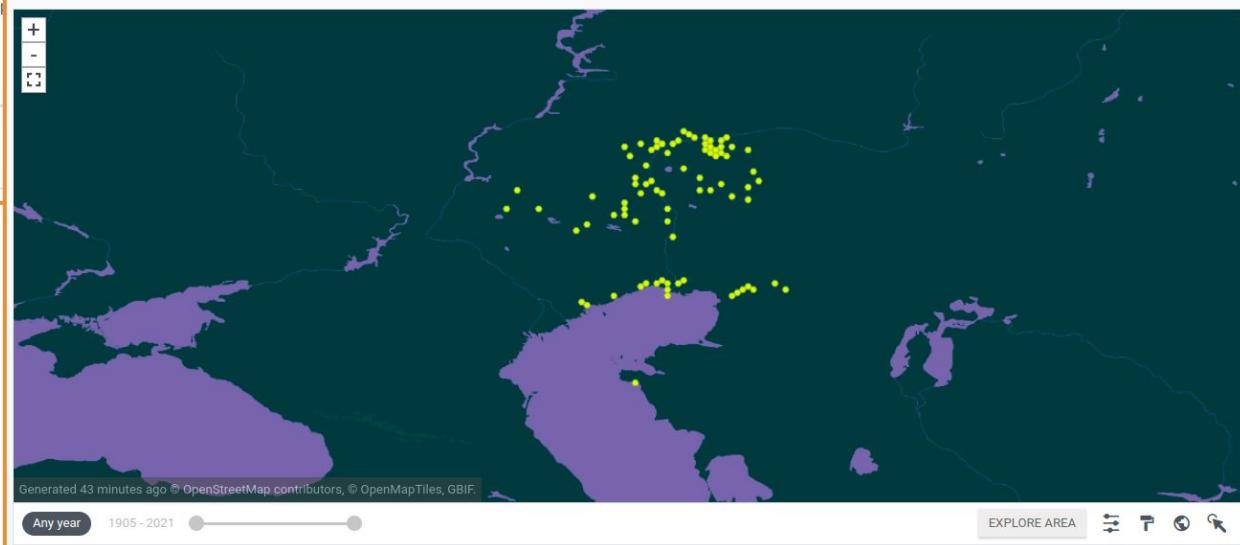
The presented data are the result of the generalization and reconciliation of literary, museum, and archival information on the distribution of lake frogs of the *P. ridibundus* complex in Kazakhstan, and new fieldwork in 2021-2022. Based on the collected material, a database has been compiled for the period from the end of the XX century to the present.

Publication date: October 13, 2022

 110
Occurrences

 100%
With taxon match

110 GEOFERENCED RECORDS



Литературные сведения

Данные собственных наблюдений

Элементарный объем информации

Нахodka (Occurrence) - Что, где, когда и кем было собрано (отмечено / записано / сфотографировано ...)

OCCURRENCE | 2 JULY 2020

Alisma plantago-aquatica L.

American waterplantain In English Collected in Russian Federation

Plantae > Tracheophyta > Liliopsida > Alismatales > Alismataceae > *Alisma*

<https://www.gbif.org/occurrence/3004106490>

Occurrence

Term	
Disposition	in collection
Associated media	https://plant.dep...mg/0.jpg
Catalogue number	MW0952835
Occurrence ID	MW0952835
Occurrence status	present
Preparations	herbarium specimen
Recorded by	N. Tihomirov



Herbarium MW

150 mm

Alisma plantago-aquatica L.
Якутия, Усть-Алданский р-н, правый берег р. Лена,
15 км от с. Усть-Алдан, 2020 г.
Листья и стебли с корнями, собранные в
день сбора гербарного экземпляра

62.5219°N 130.0419°E

02.VII.2020

Н. Тихомиров, Н. Колесов

2020

Location

Term

Country or area	Russia
Country code	RU
County	Ust'-Aldanskii raion
Decimal latitude	62.52519
Decimal longitude	130.04106
Geodetic datum	WGS84
Georeference remarks	no label data
Georeference verification status	manual verification
Georeferenced by	Kollektor
Higher geography	Siberia Yakutia Russia Sakha Ust'-Aldanskii raion



SPECIES OCCURRENCE RECORDS WITH MULTIMEDIA EVIDENCE

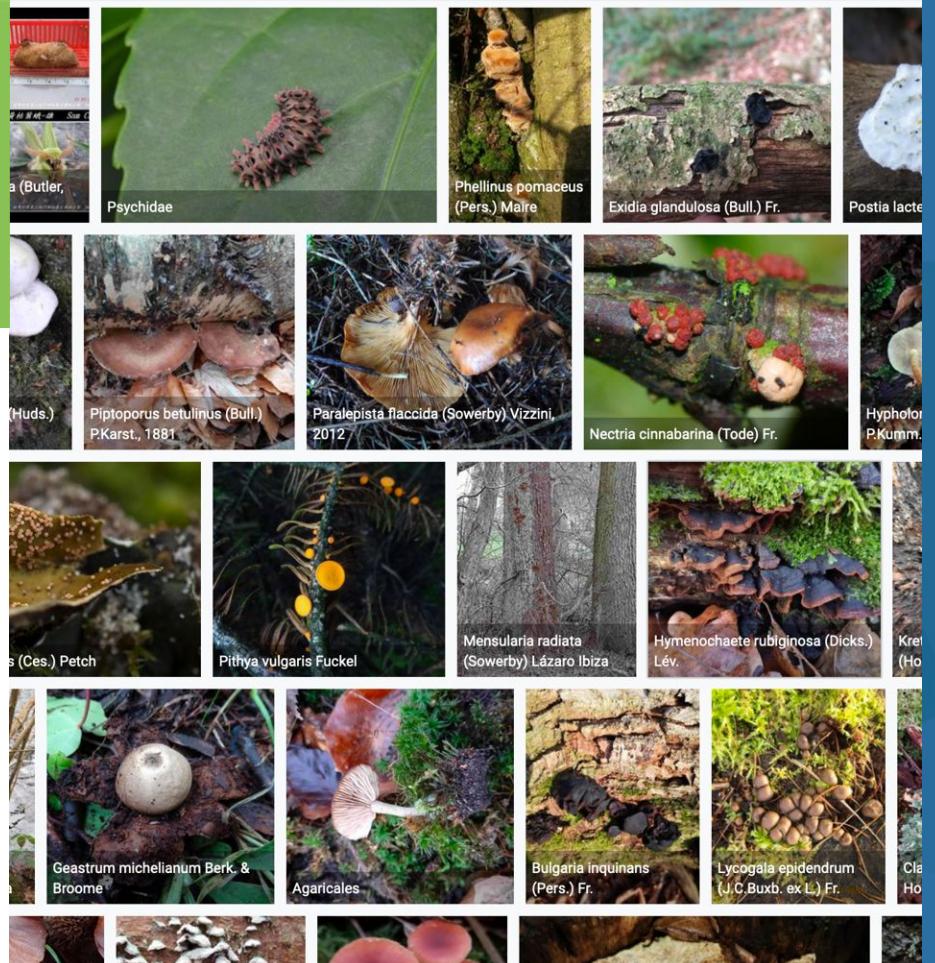
30 Sept 2022

122 million records with taxonomically identified images

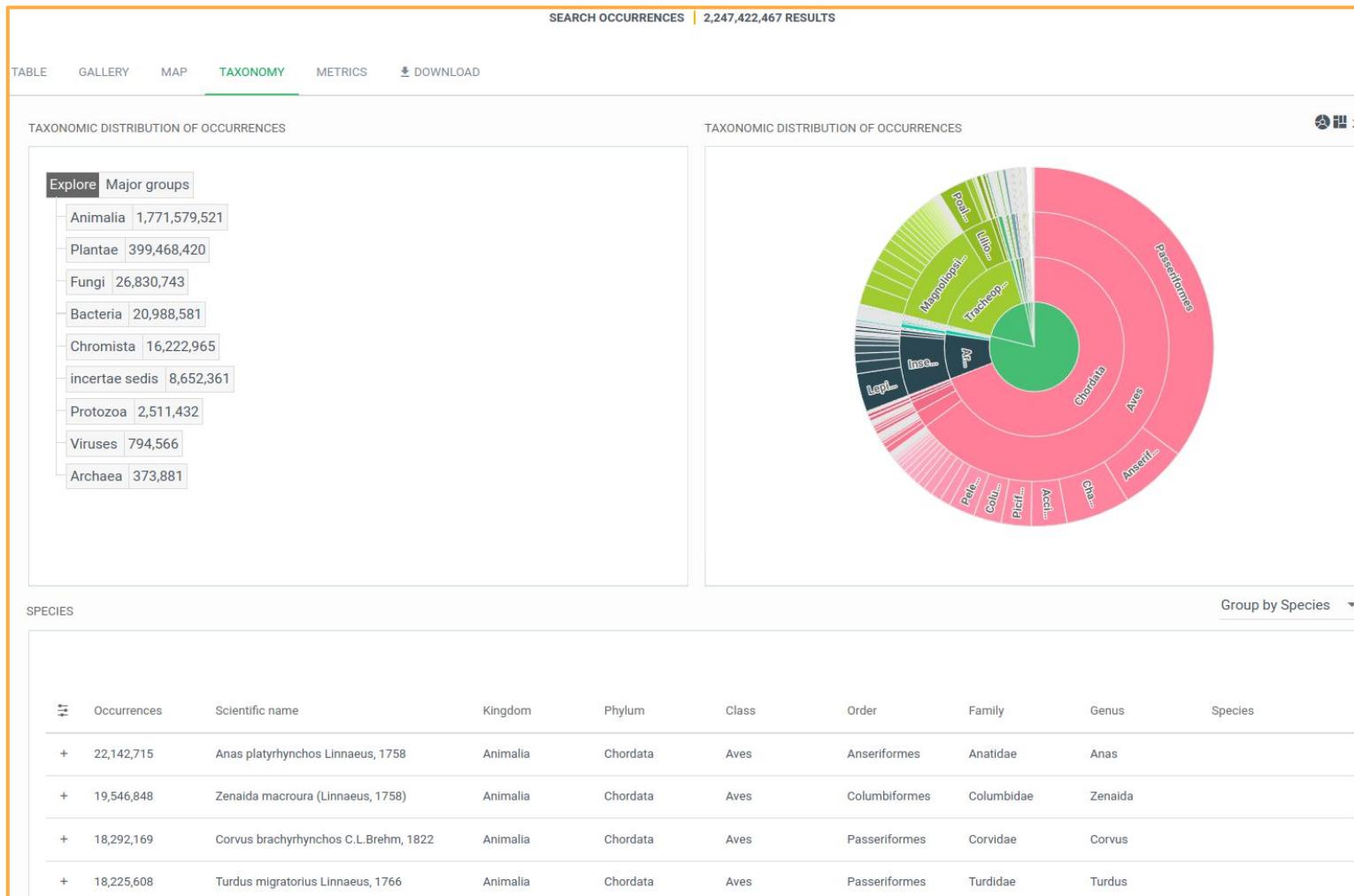
- 70 million human observations
- 48.3 million specimens
- 2.3 million material samples
- 1.4 million fossil specimens

961,740 audio files

3,599 videos



Что?



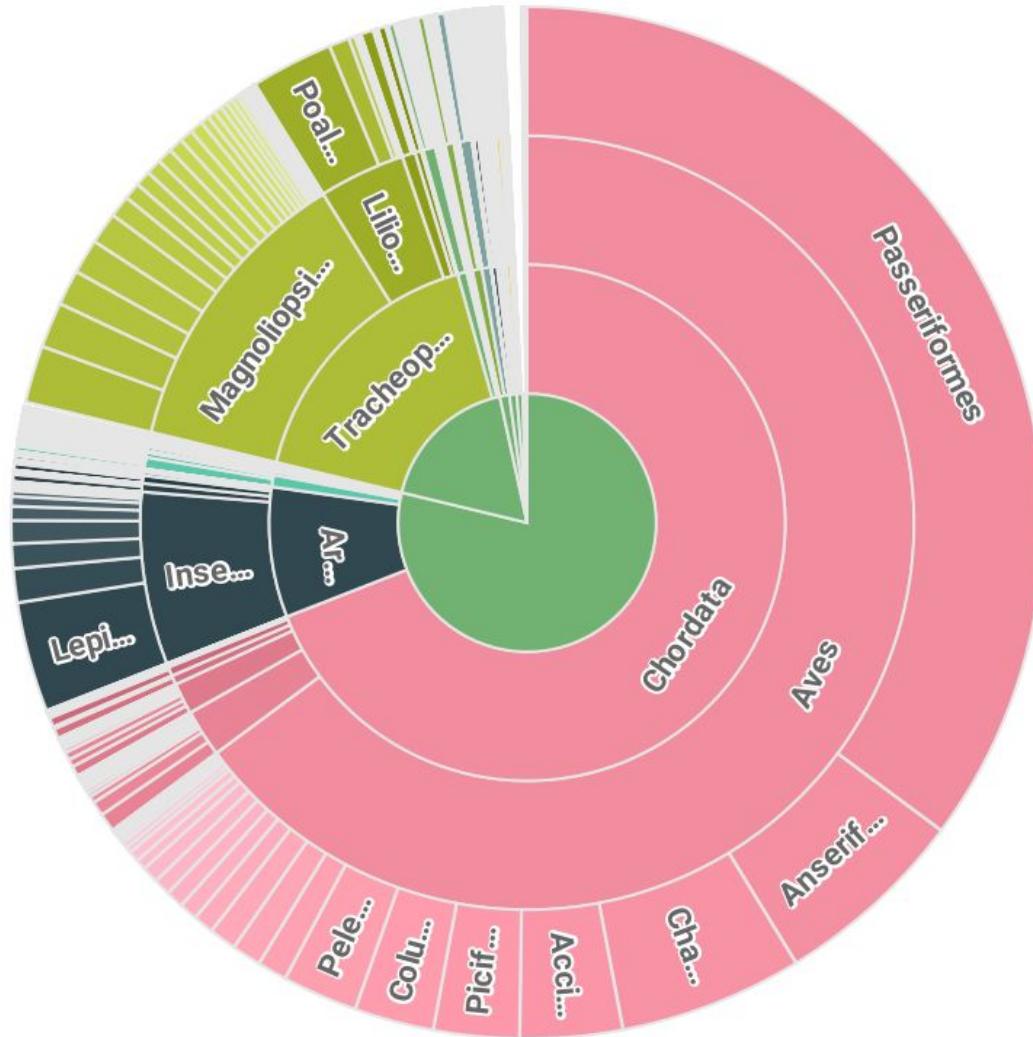
Что ?

три основные группы:

Птицы

Сосудистые растения

Насекомые



GBIF Taxonomy Backbone

CHECKLIST DATASET | REGISTERED MARCH 2, 2011

GBIF Backbone Taxonomy

Published by [GBIF Secretariat](#)

[DATASET](#)

[TAXONOMY](#)

[CONSTITUENTS](#)

[METRICS](#)

[DOWNLOAD](#)

[HOME PAGE](#)

6,783,300 RECORDS

78 CITATIONS

The GBIF Backbone Taxonomy is a single, synthetic management classification with the goal of covering all names GBIF is dealing with. It's the taxonomic backbone that allows GBIF to integrate name based information from different resources, no matter if these are occurrence datasets, species pages, names from nomenclators or external sources like EOL, Genbank or IUCN. This backbone allows taxonomic search, browse and reporting operations across all those resources in a consistent way an... [More](#)



Publication date: November 26, 2021

Metadata last modified: December 9, 2021

Hosted by: [GBIF Secretariat](#)

Licence: [CC BY 4.0](#)

 [How to cite](#)  [DOI](#) [10.15468/39omei](#)



главный Checklist и, наверно, самый большой из всех существующих 100 источников (таксономических списков)
6 783 300 записей

GBIF Taxonomy Backbone

Catalogue of Life



3 960 745 записей



Integrated
Taxonomic
Information
System
1996



Barcode Index Numbers (BINs) -
562985 записей

Systema Dipterorum

The BioSystematic Database of World Diptera



251 595 записей



223 461 записей



The Paleobiology Database
213 229 записей

Species Matching



Frank Bisby (1945 - 2011)

BIOINFORMATICS FOR BIODIVERSITY
VIEWPOINT

The Quiet Revolution: Biodiversity Informatics and the Internet

Frank A. Bisby

The massive development of biodiversity-related information systems on the Internet has created much that appears exciting but chaotic, a diversity to match biodiversity itself. This richness and the arrays of new sources are counterbalanced by the maddening difficulty in knowing what is where, or of comparing like with like. But quietly, behind the first waves of exuberance, biologists and computer scientists have started to pull together in a rising tide of coherence and organization. The fledgling field of biodiversity informatics looks set to deliver major advances that could turn the Internet into a giant global biodiversity information system.

Australia (7) and by the European Natural History Specimen Information Network (ENHSIN) team in Europe (8).

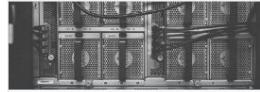
A second area for networking and interoperability is the taxonomic framework itself. Again, there are centralized models from the 1990s where organizations bring together taxonomic treatments from authors and institutions to provide a centrally collated system.

Проекты ITIS, Species 2000, European Register of Marine Species и многие другие были организованы в конце XX века.

Для разработки стандартов было создано сообщество TDWG **Taxonomic Database Working Group**. Первая конференция - 1985

TDWG - Biodiversity Information Standards

working groups

		
<p>Biodiversity Data Quality</p>	<p>Biodiversity Informatics Curriculum</p>	<p>Biodiversity Services and Clients</p>
		
<p>Biological Interactions Data</p>	<p>Citizen Science</p>	<p>Collection Descriptions</p>
		
<p>Darwin Core</p>	<p>Earth Sciences and Paleobiology</p>	<p>Genomic Biodiversity</p>
		
<p>Machine Observations</p>	<p>Observations & specimens</p>	<p>Process</p>

Conferences

		
<p>TDWG 2021 - Virtual Annual Conference</p>	<p>TDWG 2020 - A Virtual Conference</p>	<p>2019 - Leiden, The Netherlands</p>
		
<p>2018 - Dunedin, New Zealand</p>	<p>2017 - Ottawa, Canada</p>	<p>2016 - Santa Clara de San Carlos, Costa Rica</p>
		
<p>2015 - Nairobi, Kenya</p>	<p>2014 - Jönköping, Sweden</p>	<p>2013 - Florence, Italy</p>
		
<p>2012 - Beijing, China</p>	<p>2011 - New Orleans, Louisiana, USA</p>	<p>2010 - Woods Hole, Massachusetts, USA</p>

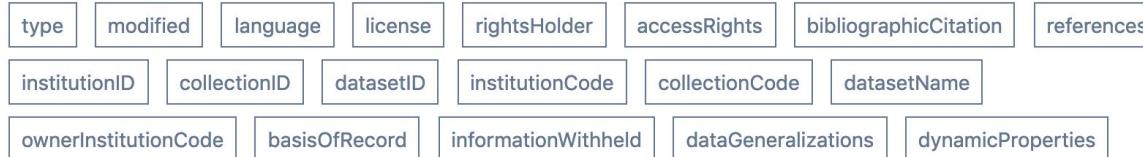
Стандарт Darwin Core (DwC)

TDWG Home Terms Guides ▾ Namespace policy

Darwin Core quick reference guide

This page provides a list of all currently recommended terms of the Darwin Core standard. Categories such as `Occurrence` or `Event` correspond to Darwin Core classes which group other terms. Convenient [files of these terms and their full history](#) can be found in the [Darwin Core repository](#).

Record-level

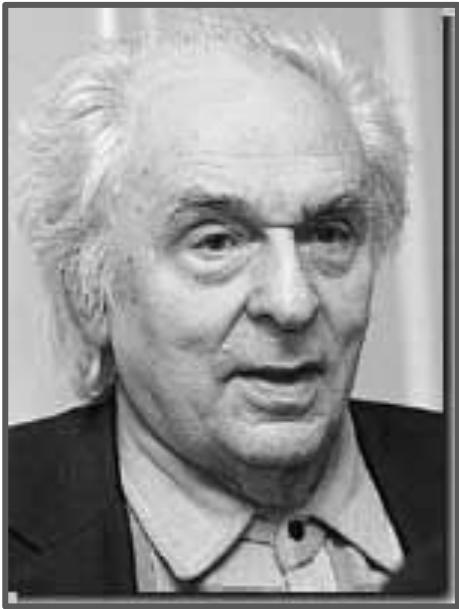


Record-level
Occurrence
Organism
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship

basisOfRecord

Property

Identifier	http://rs.tdwg.org/dwc/terms/basisOfRecord
Definition	The specific nature of the data record.
Comments	Recommended best practice is to use the standard label of one of the Darwin Core classes.
Examples	PreservedSpecimen, FossilSpecimen, LivingSpecimen, MaterialSample, Event, HumanObservation, MachineObservation, Taxon, Occurrence, MaterialCitation



Андрей Львович Лобанов 1940-2020



Большая электронно-счётная
машина - 6 (с 1968 года)
1 млн операций в секунду,
мощность - 50 кВт



"Принципы построения определителей
насекомых с использованием электронных
вычислительных машин", 1984

"Серия малых" СМ-1420



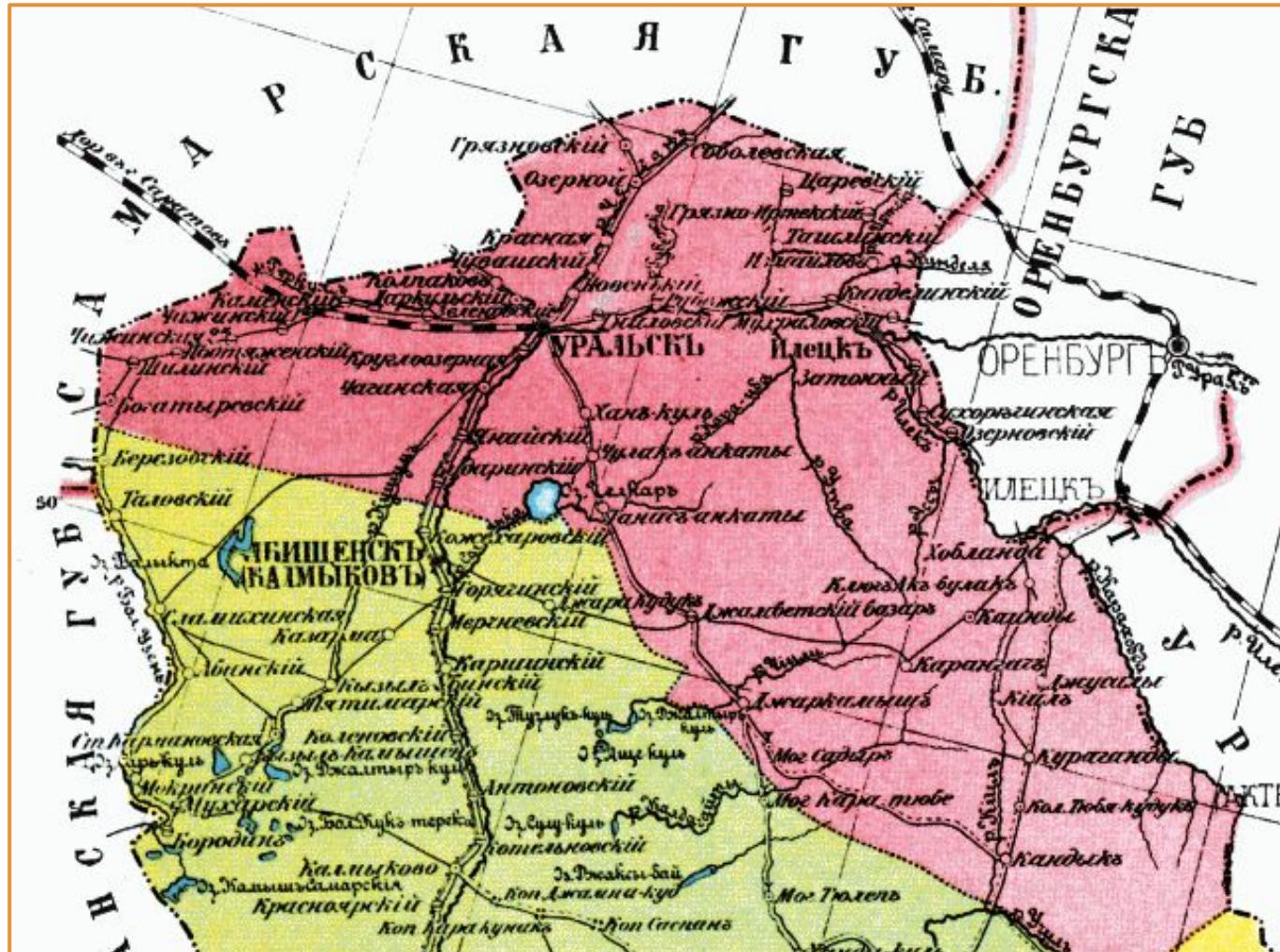
метод представления таксономических иерархических
классификаций в реляционных базах данных (концепция ZOOCOD)

Где?

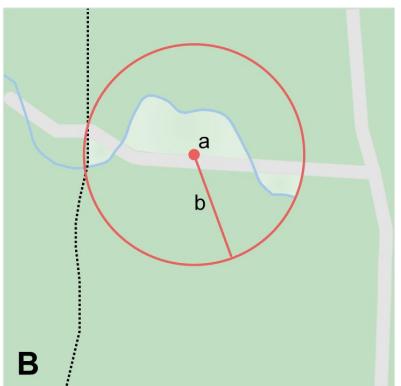
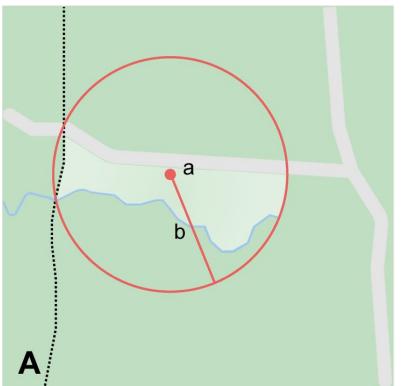
Более 90 % имеют географическую привязку

но только 745 млн (33 %) имеют оценку точности привязки

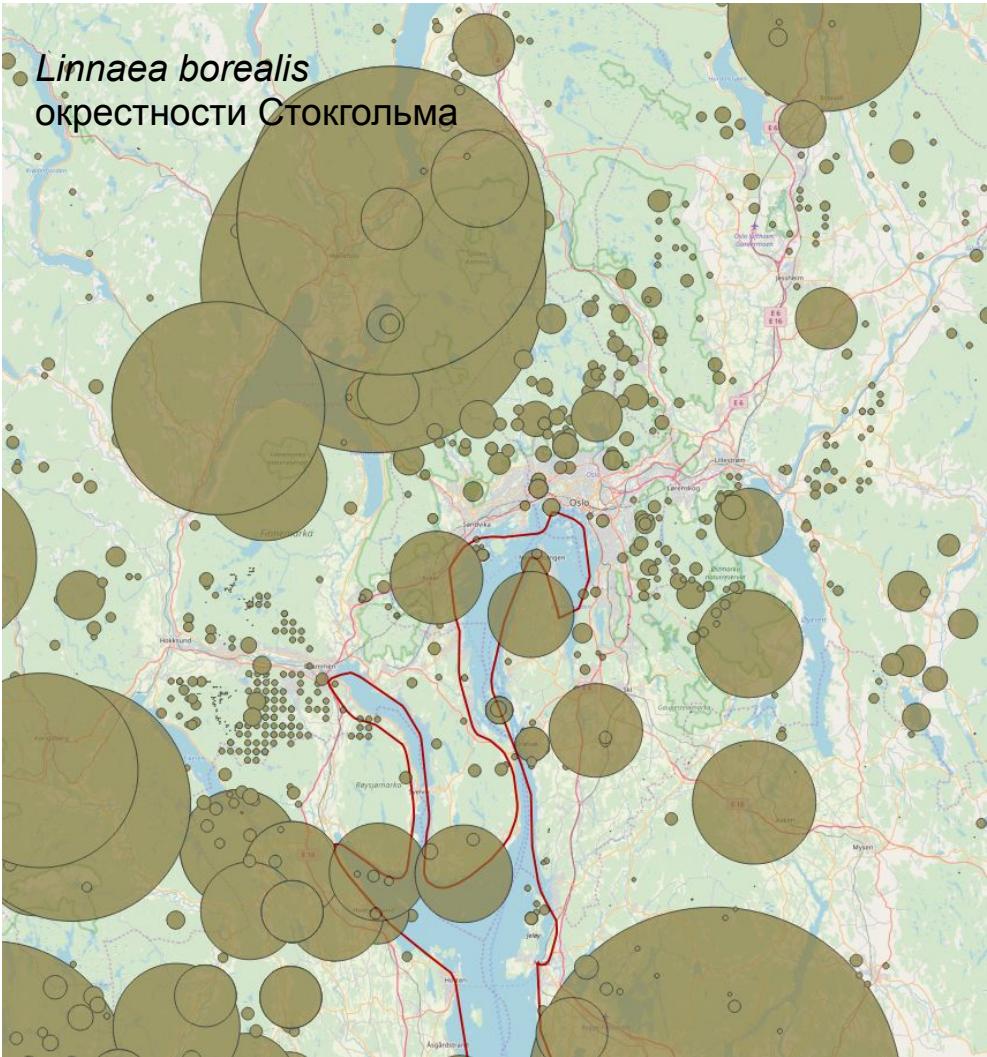
точность привязки до 100 м - менее 10%



Где? Data Quality - географическая привязка



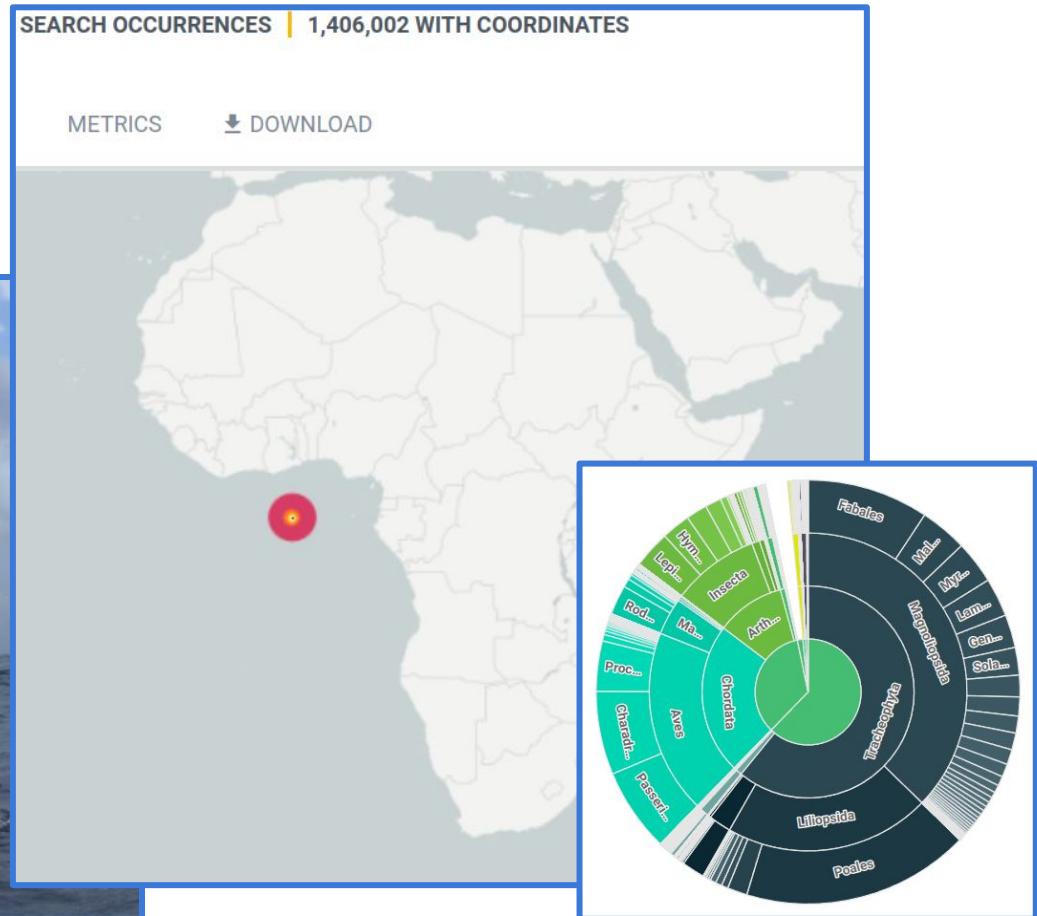
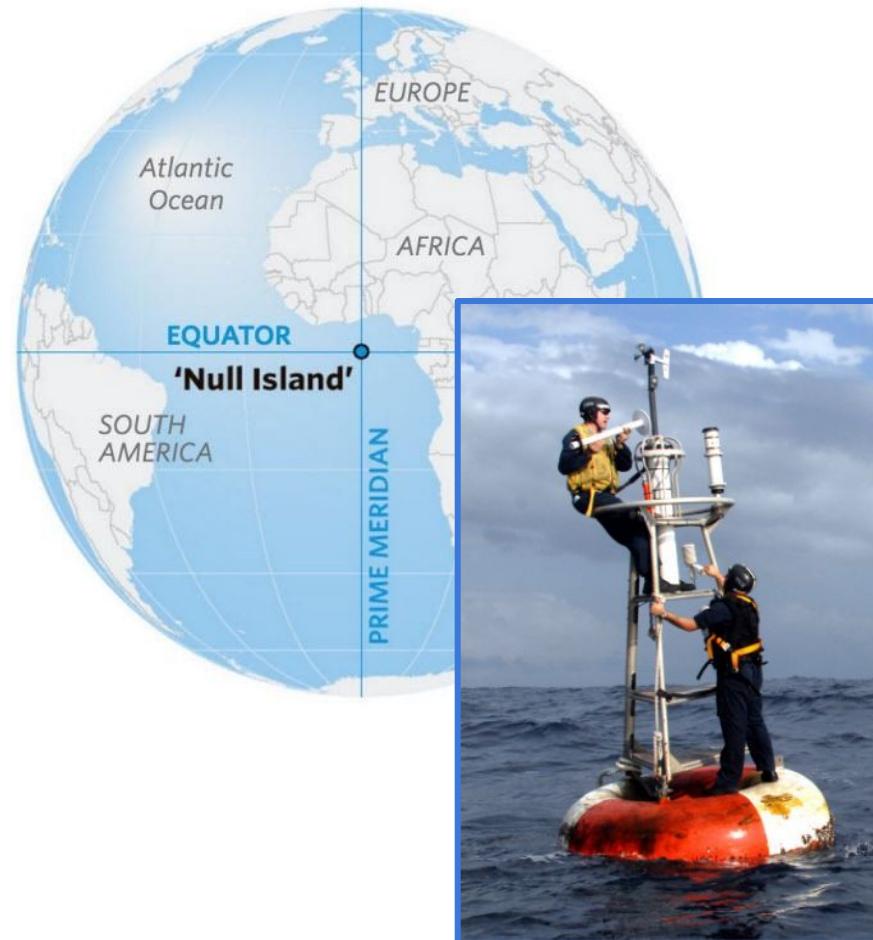
Zermoglio et al., 2020
Georeferencing Quick Reference Guide



Где?

Null Island

POINT (0,0)



Когда?

первый миллион находок
1600-1833 гг.

OCCURRENCES PER YEAR

OCCURRENCES PER YEAR

Occurrences

Click and drag in the plot area to zoom in

Occurrences

OCCURRENCES PER YEAR

Click and drag in the plot area to zoom in

750k

500k

250k

0

1600

1650

1700

1750

1800

1850

1900

1950

2000

Occurrences

1900

1950

2000

2050

2100

2150

2200

2250

2300

2350

2400

2450

2500

2550

2600

2650

2700

2750

2800

2850

2900

2950

3000

141,940,191 other or unknown

Occurrences

0

1600

1650

1700

1750

1800

1850

1900

1950

2000

2050

2100

2150

2200

2250

2300

2350

2400

2450

2500

2550

2600

2650

2700

2750

2800

2850

2900

2950

3000

3050

3100

3150

3200

3250

3300

3350

3400

3450

3500

3550

3600

3650

3700

3750

3800

3850

3900

3950

4000

4050

4100

4150

4200

4250

4300

4350

4400

4450

4500

4550

4600

4650

4700

4750

4800

4850

4900

4950

5000

5050

5100

5150

5200

5250

5300

5350

5400

5450

5500

5550

5600

5650

5700

5750

5800

5850

5900

5950

6000

6050

6100

6150

6200

6250

6300

6350

6400

6450

6500

6550

6600

6650

6700

6750

6800

6850

6900

6950

7000

7050

7100

7150

7200

7250

7300

7350

7400

7450

7500

7550

7600

7650

7700

7750

7800

7850

7900

7950

8000

8050

8100

8150

8200

8250

8300

8350

8400

8450

8500

8550

8600

8650

8700

8750

8800

8850

8900

8950

9000

9050

9100

9150

9200

9250

9300

9350

9400

9450

9500

9550

9600

9650

9700

9750

9800

9850

9900

9950

10000

10050

10100

10150

10200

10250

10300

10350

10400

10450

10500

10550

10600

10650

10700

10750

10800

10850

10900

10950

11000

11050

11100

11150

11200

11250

11300

11350

11400

11450

11500

11550

11600

11650

11700

11750

11800

11850

11900

11950

12000

12050

12100

12150

12200

12250

12300

12350

12400

12450

12500

12550

12600

12650

12700

12750

12800

12850

12900

12950

13000

13050

13100

13150

13200

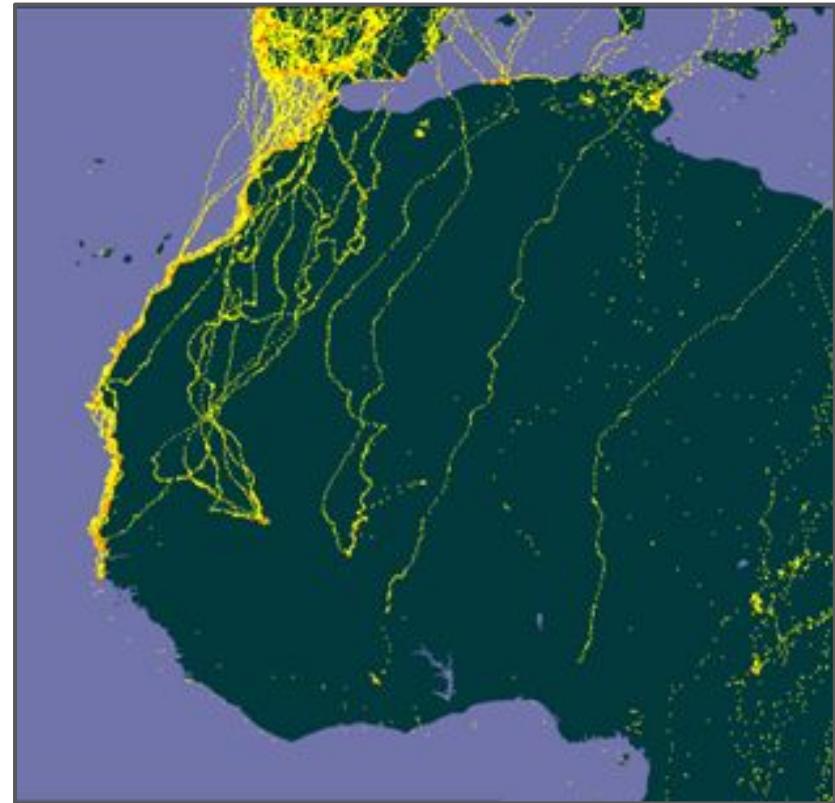
13250

13300

13350

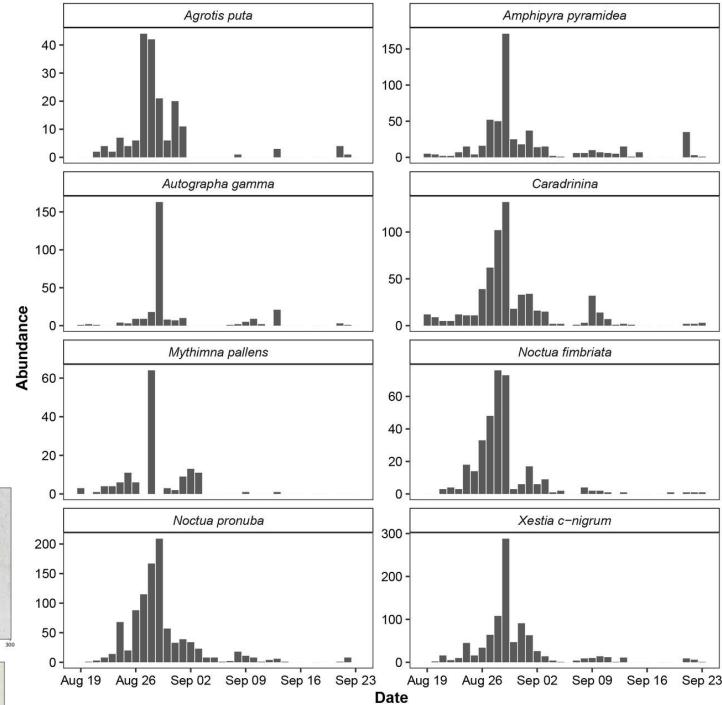
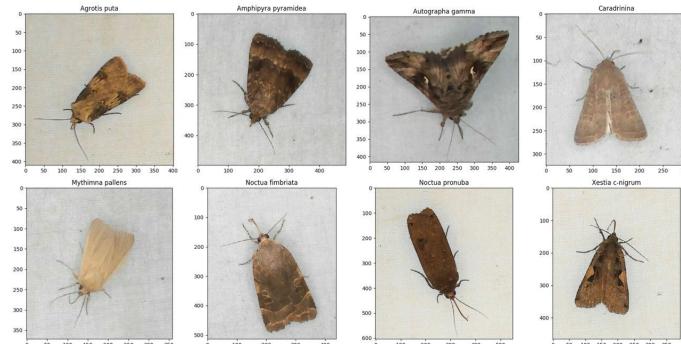
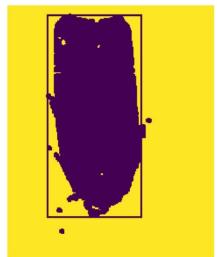
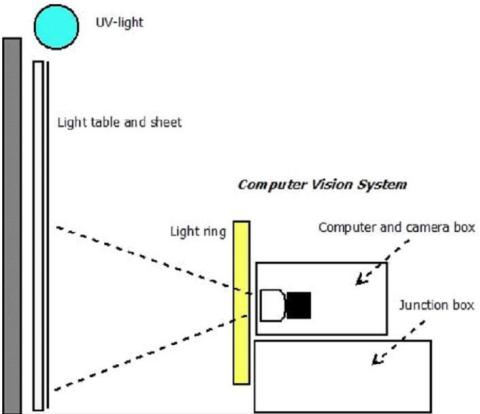
13400

Типы данных - автоматические наблюдения



GPS-трекеры на крупных перелетных птицах

Автоматизированный учёт насекомых



10.3390/s21020343

Точки отсутствия

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

Occurrences



Search all fields

[Simple](#)[Advanced](#)

Occurrence status

[Everything](#)[Present](#)[Absent](#)

'Absent' is applied to an occurrence record when a survey of a taxon at a specific time and place encounters no specimens

Licence

Scientific name

Verbatim scientific name

SEARCH OCCURRENCES | 28,276,367 RESULTS

[TABLE](#)[GALLERY](#)[MAP](#)[TAXONOMY](#)[METRICS](#)[DOWNLOAD](#)[Scientific name](#)[Country or area](#)[Coordinates](#)[Month & year](#)

Tachybaptus ruficollis (Pallas, 1764)

Norway

59.1N, 9.6E

2022 January

Turdus pilaris Linnaeus, 1758

Norway

69.6N, 18.9E

2022 January

Fringilla montifringilla Linnaeus, 1758

Norway

63.6N, 10.7E

2022 January

Periparus ater (Linnaeus, 1758)

Norway

59.9N, 6.6E

2022 January

Corythornis cristatus (Pallas, 1764)

South Africa

25.1S, 27.8E

2022 January

Dendrocopos major (Linnaeus, 1758)

Norway

62.4N, 11.0E

2022 January

Кем?

[recordedBy](#), [recordedByID](#) - собрано

[georeferencedBy](#) - привязано в пространстве

[identifiedBy](#), [identifiedByID](#) - определено

[measurementDeterminedBy](#) - измерили или оценили какие-то дополнительные значения (биомасса, размеры организма, значения факторов среды)

для каждой роли может быть указан один специалист или несколько, а также группа (лаборатория) или организация

Типы данных, публикуемых через GBIF



Specimen



Fossil



Observation
Human observation
Living specimen



Literature occurrence



Material sample



Machine observation



Метаданные - данные о данных

Название набора данных

Тип данных

Объем (число записей)

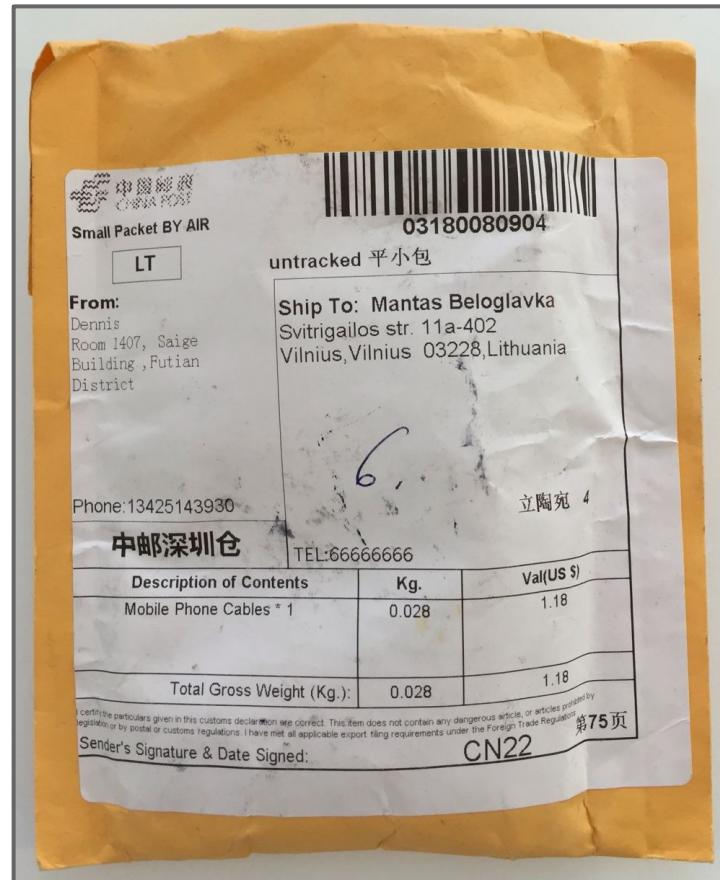
Краткое описание

Таксономические группы

Географический охват

Временной диапазон

— 10 —



Типы наборов данных

Описание набора данных

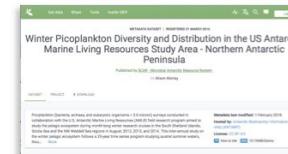
Список видов

Отдельные находки (количественные характеристики)

Характеристики сборов

Metadata only dataset

No data content required.
You know what is in your collection and you can describe its content and scope but you cannot make the data content available on GBIF.



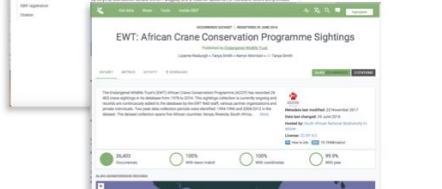
Checklist

- Scientific names of organisms sharing a common theme or feature (for example: medicinal use).



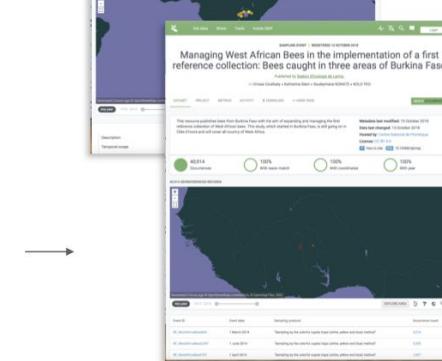
Occurrence Dataset

- Scientific names of organisms observed or specimens collected,
- Observation or sampling date (year),
- Observation or sampling location (at least country).

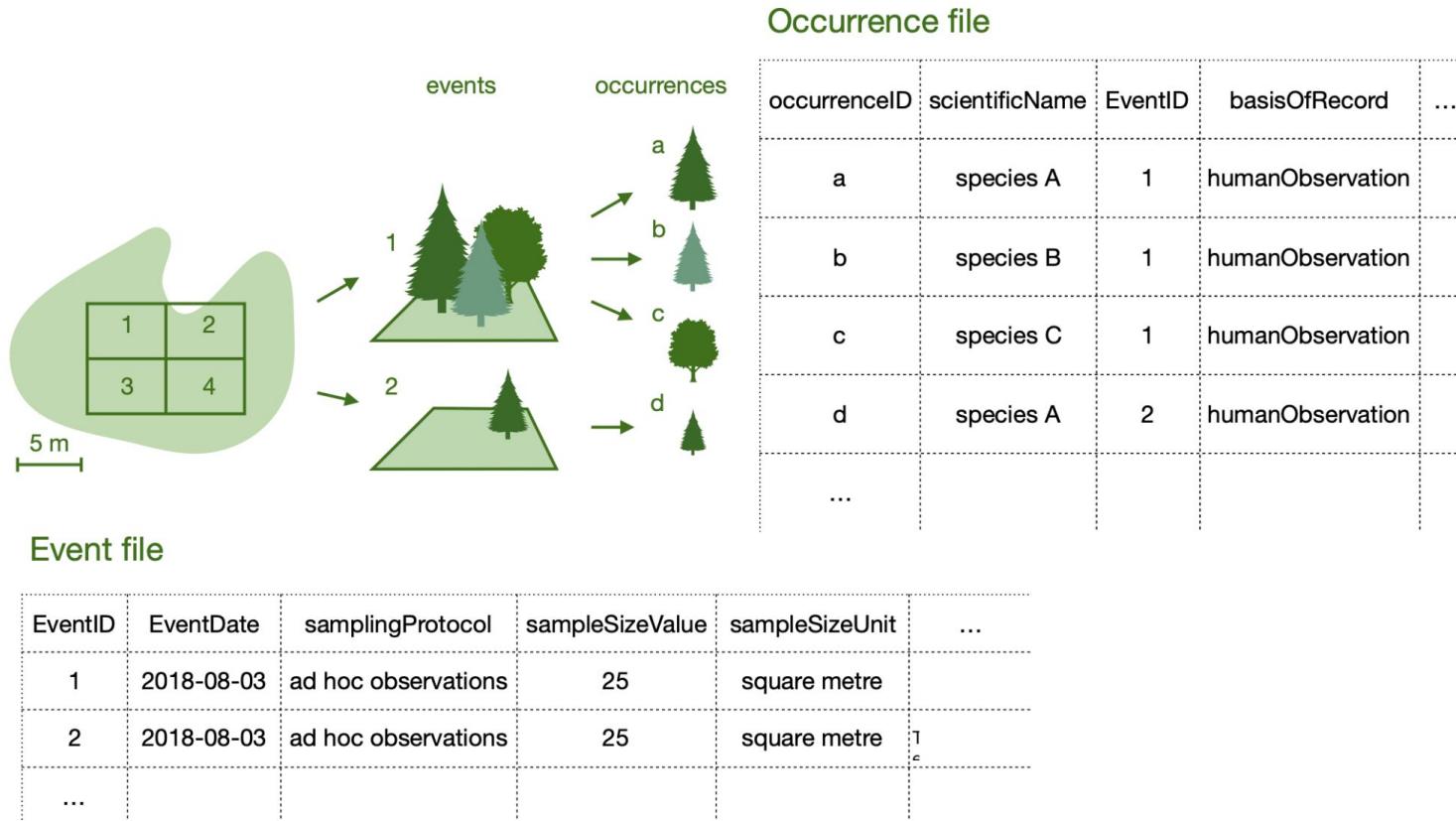


Sampling-Event Dataset

- Scientific names of organisms observed or specimens collected,
- Sampling date,
- Observation or sampling location,
- Sampling protocol.



Sampling Event dataset



<https://data-blog.gbif.org/post/choose-dataset-type>

Подробность и варианты Occurrence dataset

- Присутствие или отсутствие вида
- Число экземпляров вида в данной точке
- Если есть характеристики каждого экземпляра, которые отличают его от другого: онтогенетическое состояние, биомасса и проч., то в качестве находки можно указывать экземпляр
- Перемещение отдельного экземпляра в пространстве - автоматическая фиксация координат через определенные промежутки времени

Источники данных

OCCURRENCES PER DATASET



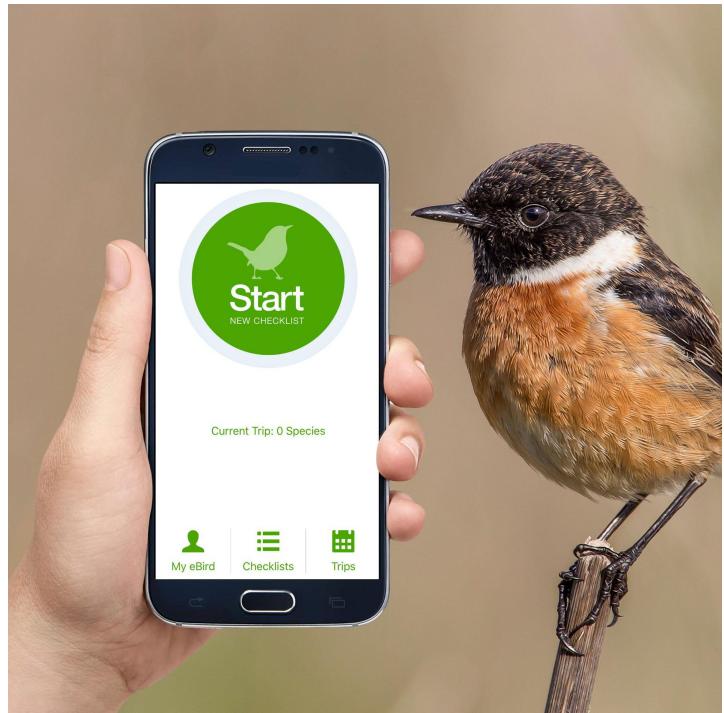
eBird

Всемирная сеть
наблюдений за птицами
> 1 000 000 000 находок

<https://ebird.org/home>

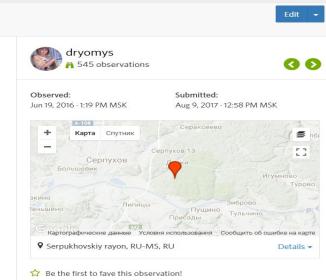


adkl786492 Kostroma Region, Russia
Kolotilin Andrian (C) GeoPhoto.Ru

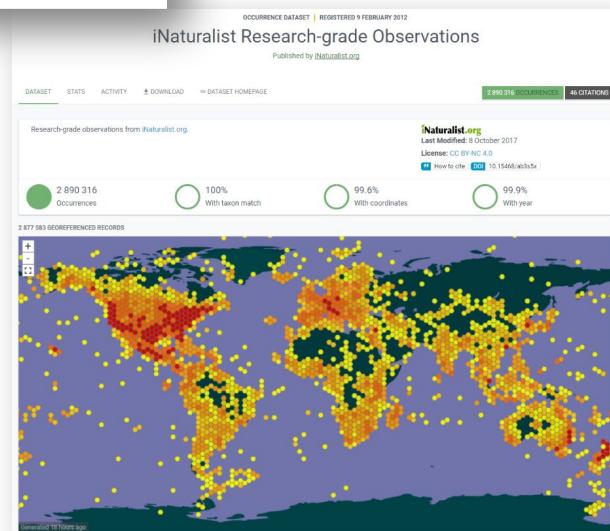


Сообщество натуралистов – система [Inaturalist.org](https://inaturalist.org)

~ 74 млн Research Grade наблюдений



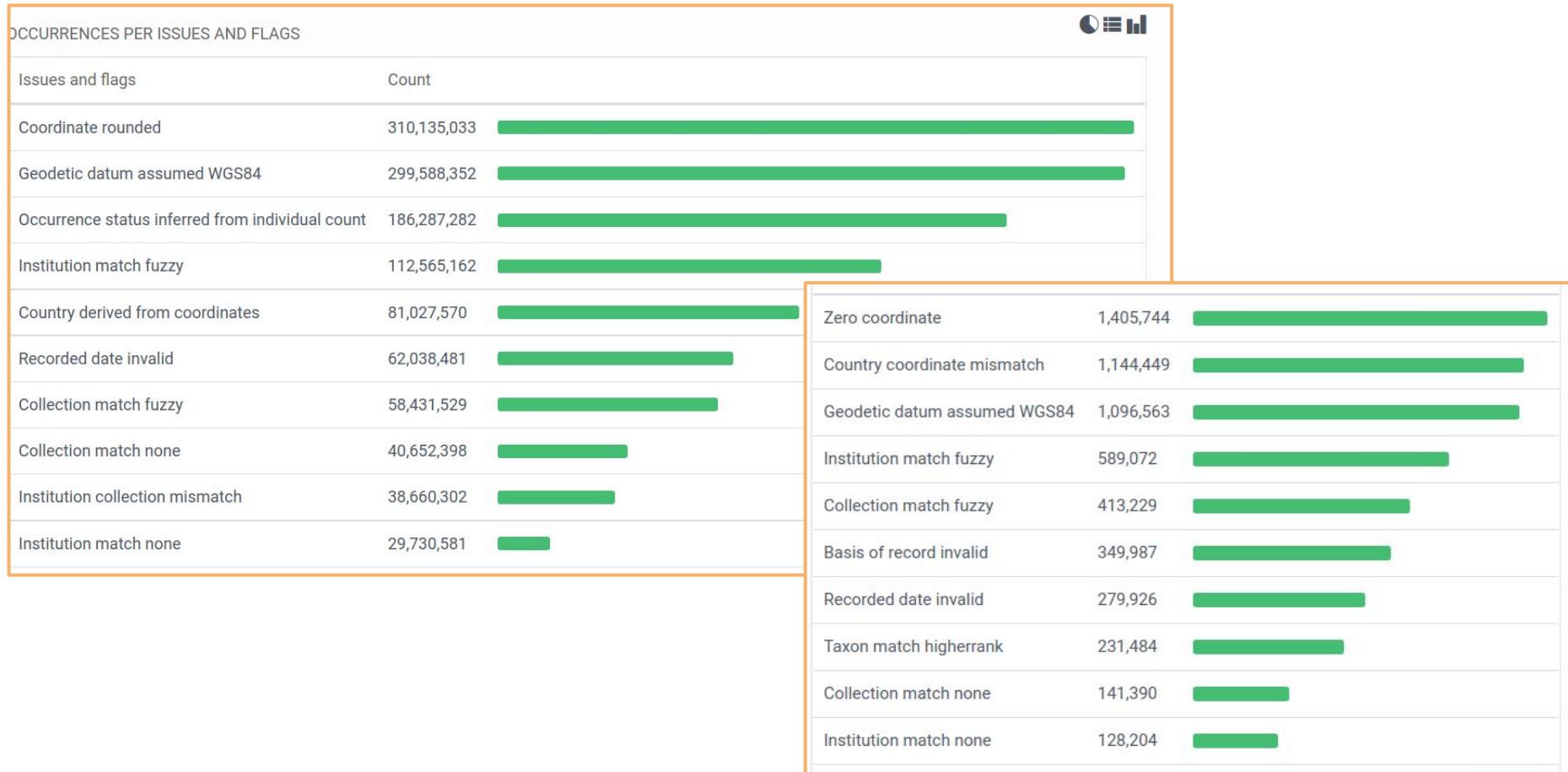
> 52 млн. записей



~ 2.5 млн. участников
> 120 млн наблюдений



Качество данных - Data Quality



Occurrence-only datasets

Learn more about occurrence-only and other classes of datasets currently supported

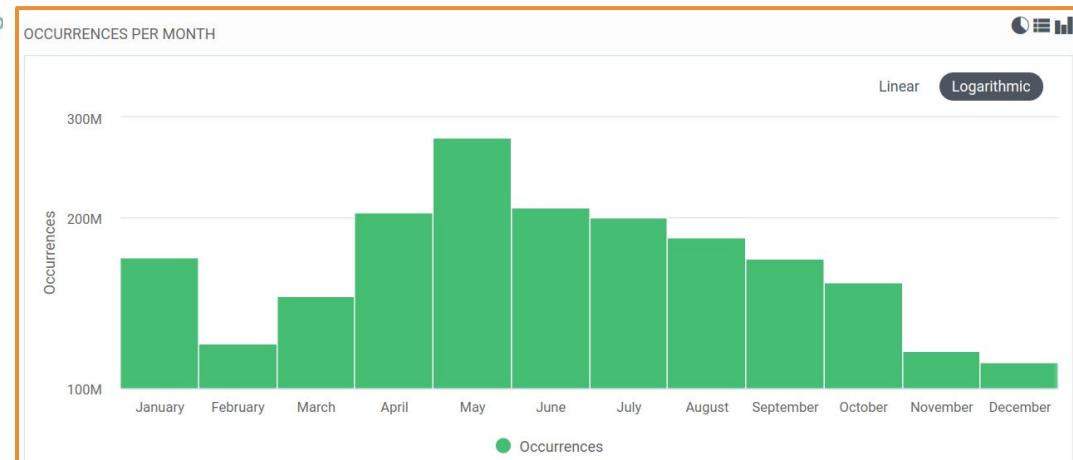
Качество данных - Data Quality

Darwin Core record details

Term	Status
occurrenceID	Required
basisOfRecord	Required
scientificName	Required
eventDate	Required
countryCode	Required

taxonRank	Strongly recommended
kingdom	Strongly recommended
decimalLatitude & decimalLongitude	Strongly recommended
geodeticDatum	Strongly recommended
coordinateUncertaintyInMeters	Strongly recommended
individualCount, organismQuantity & organismQuantityType	Strongly recommended

informationWithheld	Share if available
dataGeneralizations	Share if available
eventTime	Share if available
country	Share if available

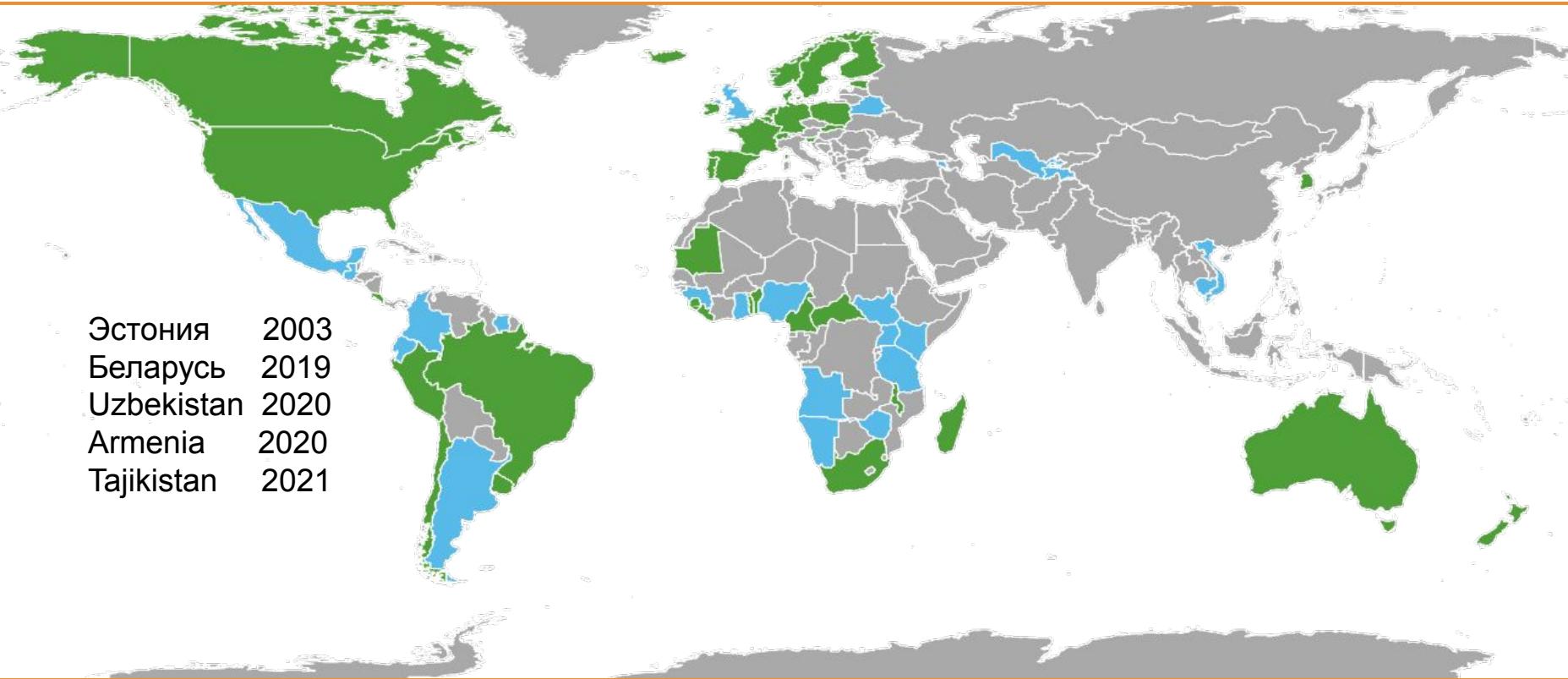


Checklist datasets

taxonID	Required
scientificName	Required
taxonRank	Required
kingdom	Strongly recommended
parentNameUsageID	Strongly recommended
acceptedNameUsageID	Strongly recommended
vernacularName	Share if available

<https://www.gbif.org/data-quality-requirements>

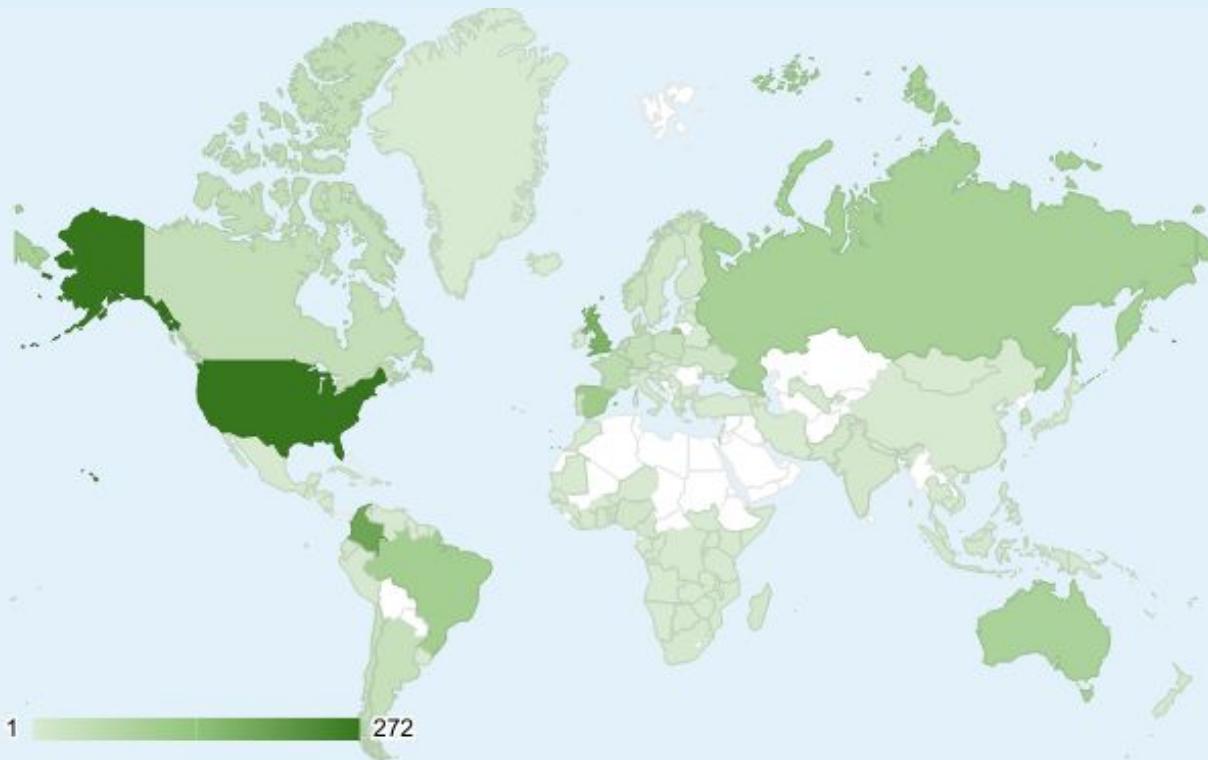
GBIF - не только портал



40 стран-участниц 23 ассоциированных страны 43 прочих
участника 1985 публикующих организаций

GBIF NETWORK OF DATA PUBLISHING INSTITUTIONS

30 September 2022



132

countries/territories
with institutions
sharing data
through GBIF

Top 10 countries: number of data publishers

1	United States	272
2	Colombia	189
3	United Kingdom	164
4	Spain	112
5	Brazil	102
6	Russian Federation	95
7	Australia	88
8	France	56
9	Canada	44
10	Netherlands	41

Российские участники GBIF

на начало 2019 года



Новости портала GBIF

Вебинары

Учебные курсы

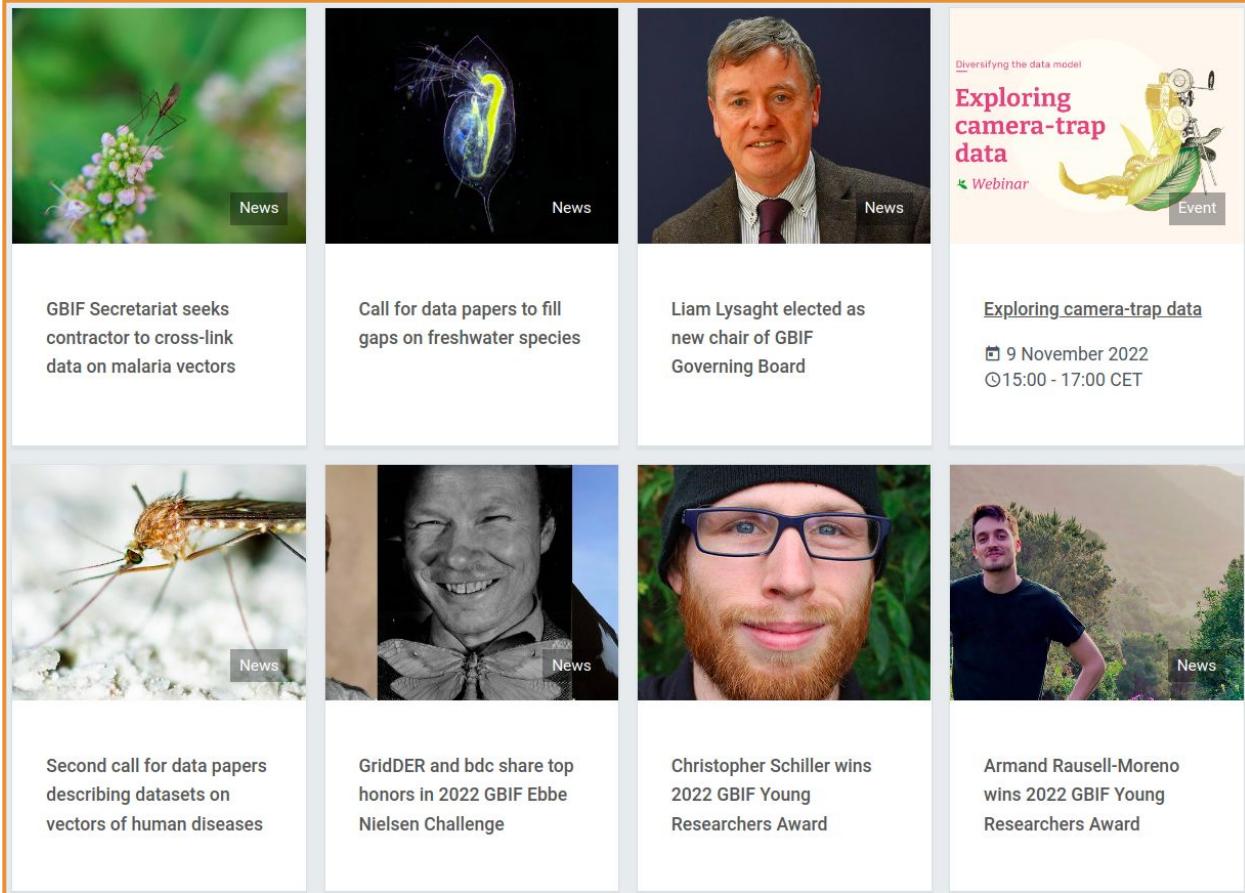
Конкурсы поддержки проектов

Премии для молодых учёных

[Data papers:](#) 8 ноября
[Bringing data to light](#)

9 ноября

[Exploring camera-trap data](#)



The grid displays eight news items:

- GBIF Secretariat seeks contractor to cross-link data on malaria vectors** (Thumbnail: a mosquito on a flower)
- Call for data papers to fill gaps on freshwater species** (Thumbnail: a freshwater species, possibly a fish larva)
- Liam Lysaght elected as new chair of GBIF Governing Board** (Thumbnail: Liam Lysaght)
- Exploring camera-trap data** (Thumbnail: a camera trap with a green leaf)
- Second call for data papers describing datasets on vectors of human diseases** (Thumbnail: a mosquito)
- GridDER and bdc share top honors in 2022 GBIF Ebbe Nielsen Challenge** (Thumbnail: a smiling man with a butterfly)
- Christopher Schiller wins 2022 GBIF Young Researchers Award** (Thumbnail: a man with a beard and glasses)
- Armand Rausell-Moreno wins 2022 GBIF Young Researchers Award** (Thumbnail: a man outdoors)