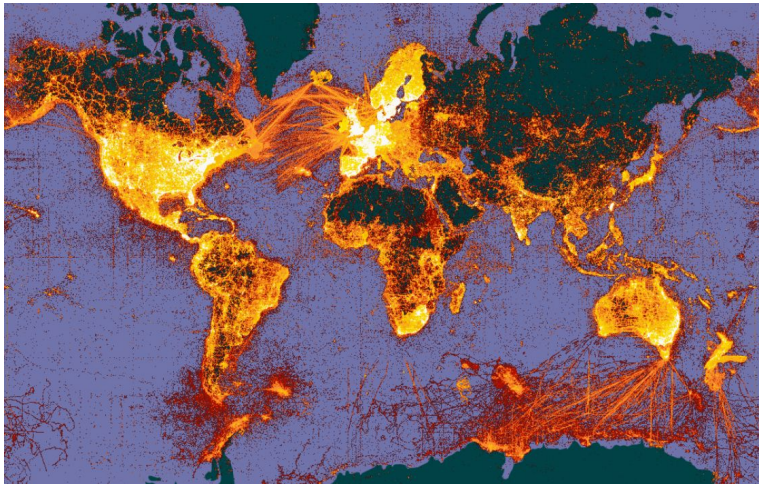
Биоәртүрлілікті зерттеудегі цифрлық технологиялар
Digital technologies in biodiversity research
Цифровые технологии в исследовании биоразнообразия

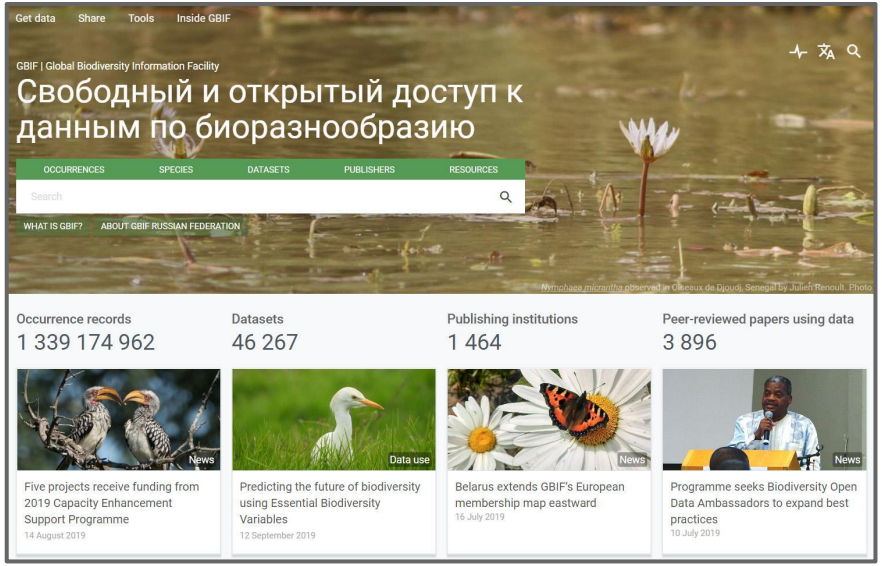
ЛЕКЦИЯ 6

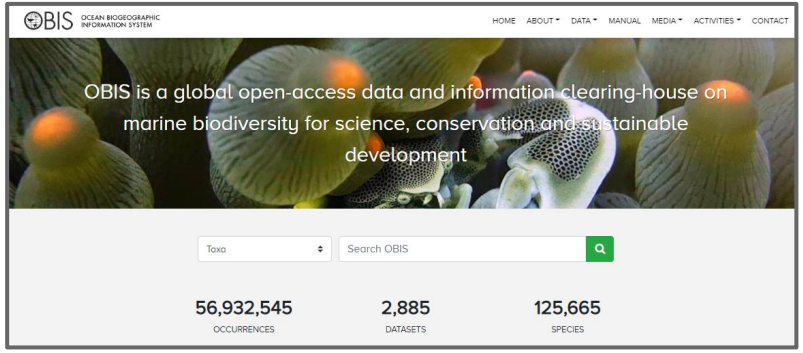
Портал GBIF - крупнейший мировой репозиторий открытых данных о биоразнообразии

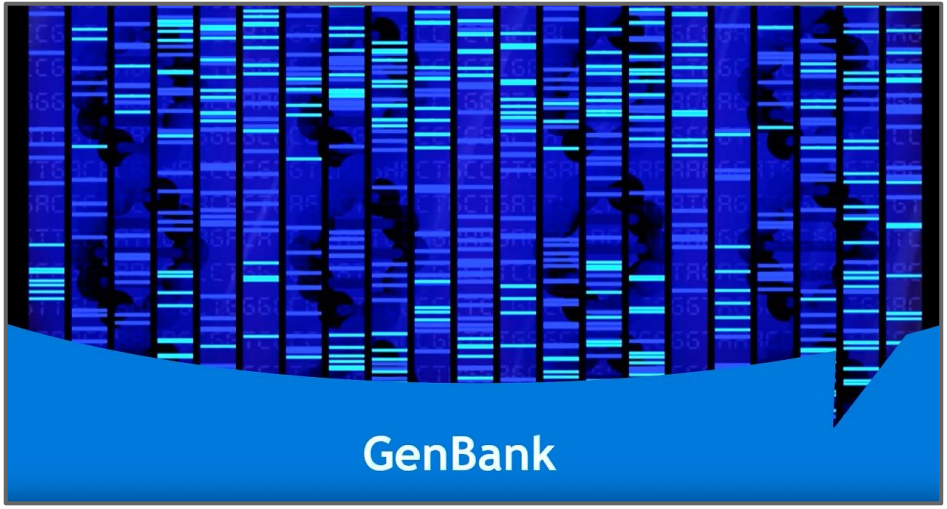


Слайды CC BY:
Dag Endresen, GBIF Norway
Наталья Иванова,
Максим Шашков

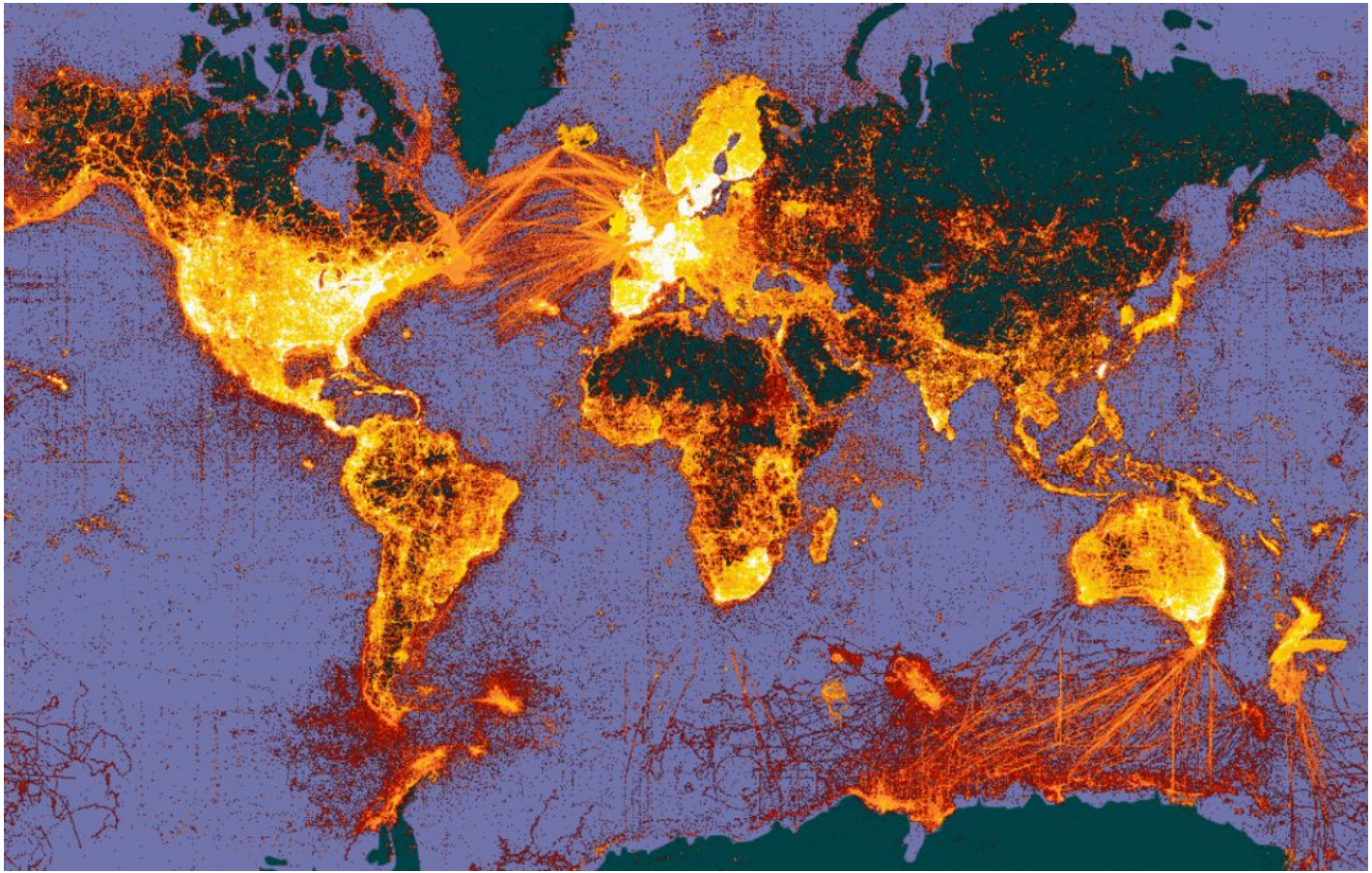
Репозитории







GBIF - крупнейший агрегатор данных о биоразнообразии



1 339 174 017
находок

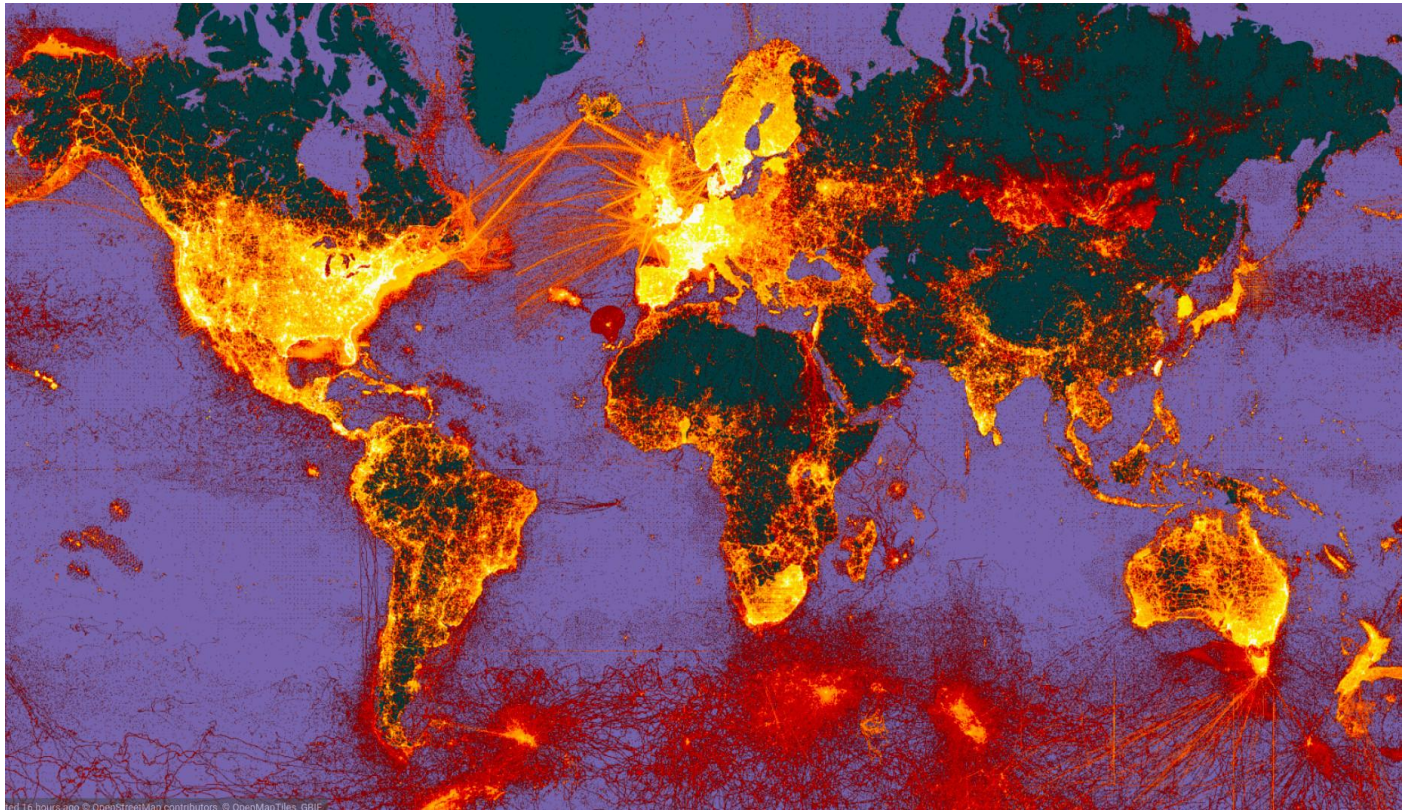
46 267
наборов данных

1464
организаций

3896 статей с использованием данных GBIF

15 сентября 2019 г.

GBIF - крупнейший агрегатор данных о биоразнообразии



2 246 485 892
находок

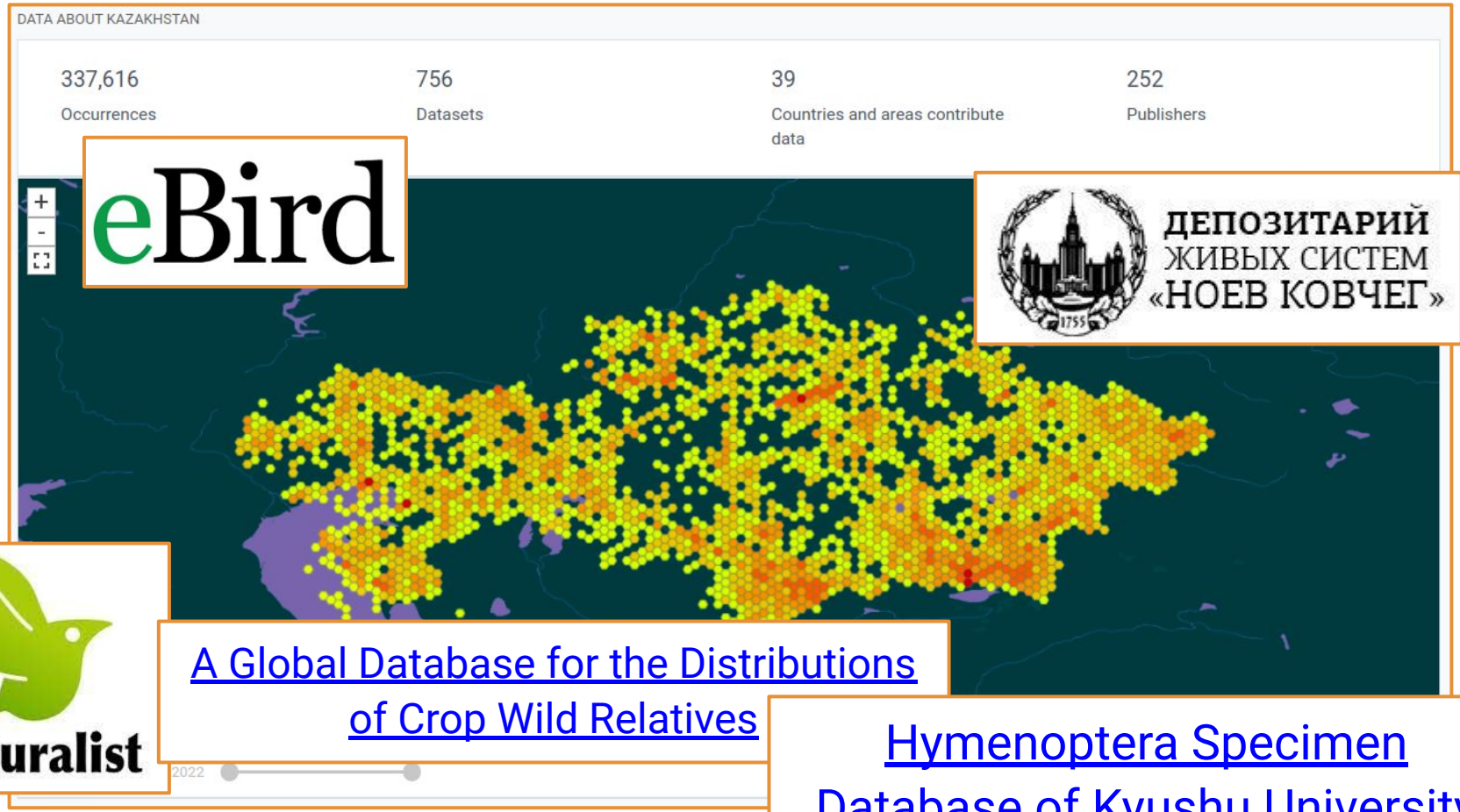
79 054
наборов данных

1926
организаций

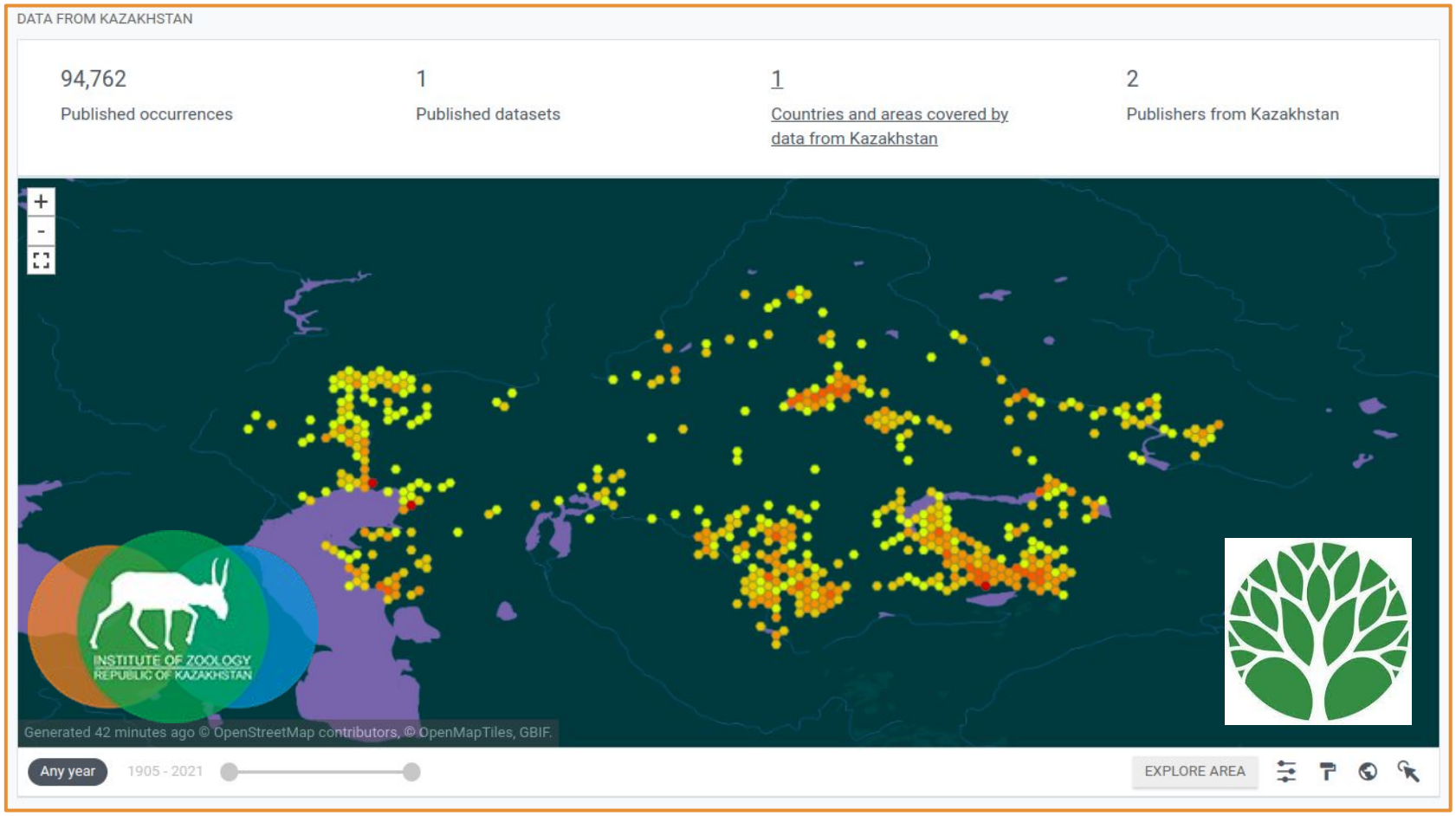
7902 статей с использованием данных GBIF

1 ноября 2021 г.

Казахстан в GBIF



Казахстан в GBIF



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

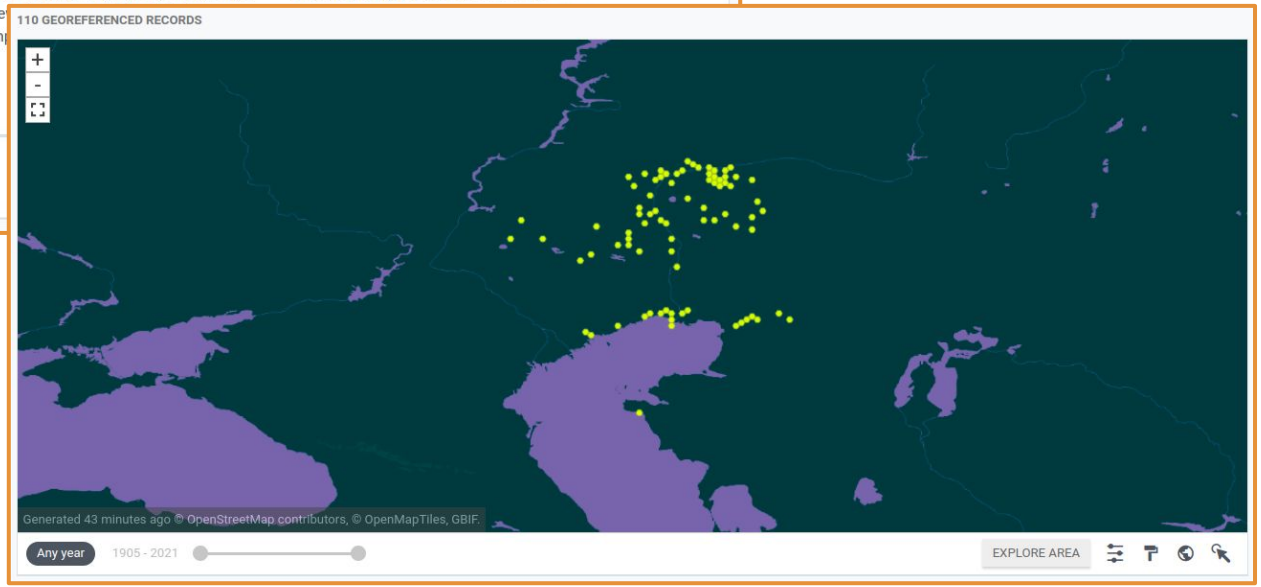
DATASET | METRICS | ACTIVITY | DOWNLOAD | HOME PAGE | 110 OCCURRENCES

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 ne fieldwork in 2021-2022. Based on the collected material, a database has been comp today for the period from the end of the XX century to the present.

Publication date: October 13, 2022

110 Occurrences

100% With taxon match



Литературные сведения

Данные собственных наблюдений

Элементарный объем информации

Находка (Occurrence) - **Что**, **где**, **когда** и **кем** было собрано (отмечено / записано / сфотографировано …)

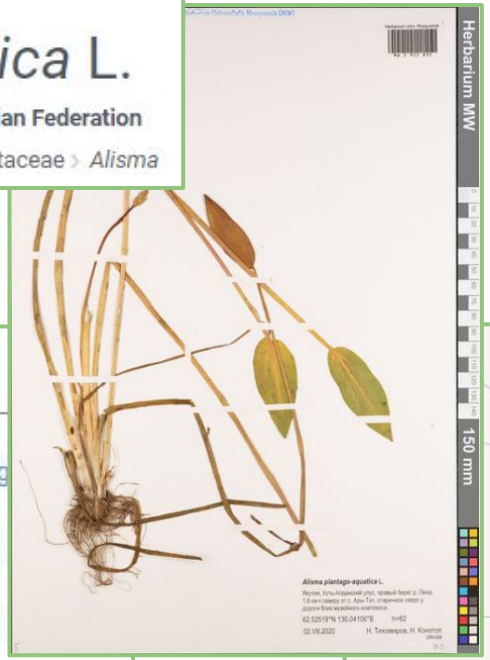
OCCURRENCE | 2 JULY 2020

Alisma plantago-aquatica L.

American waterplantain In English **Collected in Russian Federation**

Plantae › Tracheophyta › Liliopsida › Alismatales › Alismataceae › *Alisma*

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

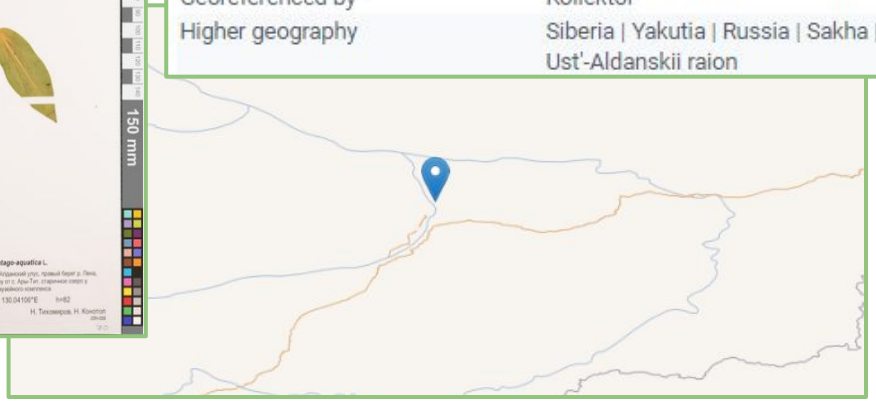


Occurrence

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

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

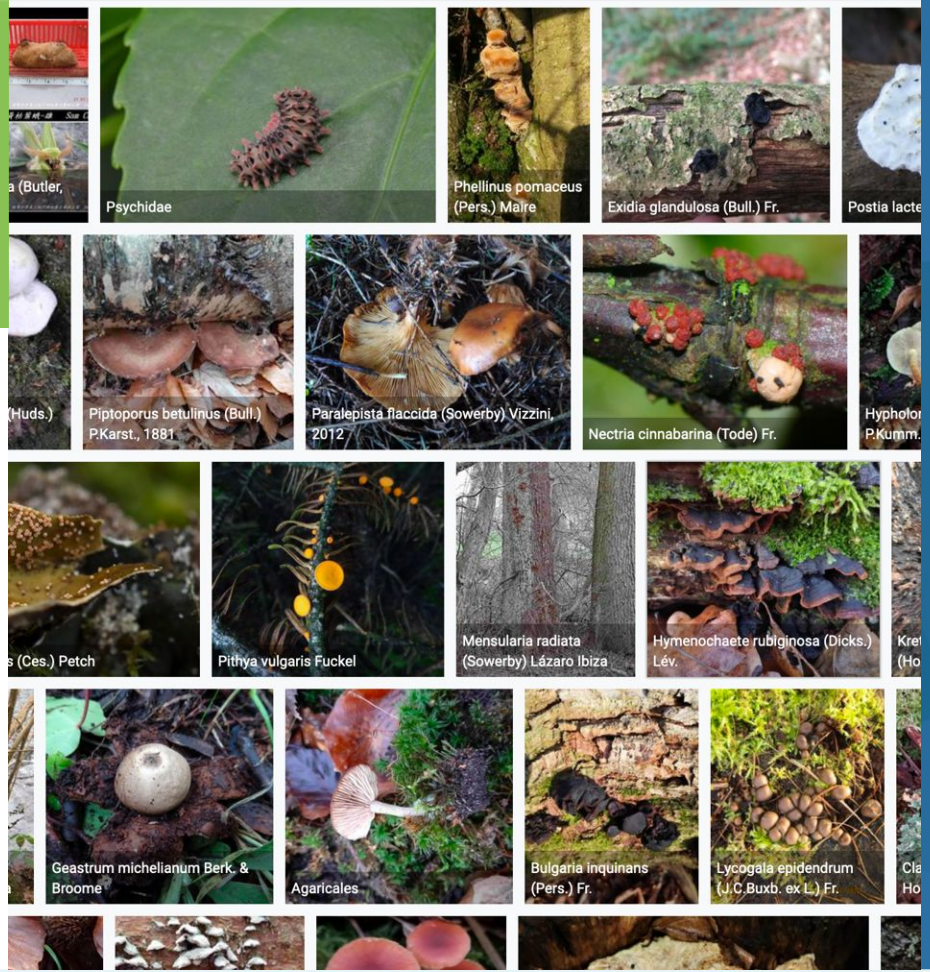
30 Sept 2022

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 ?

SEARCH OCCURRENCES | 2,247,422,467 RESULTS

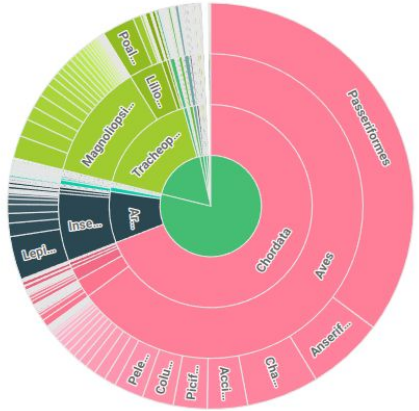
TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

TAXONOMIC DISTRIBUTION OF OCCURRENCES

Explore Major groups

- Animalia 1,771,579,521
- Plantae 399,468,420
- Fungi 26,830,743
- Bacteria 20,988,581
- Chromista 16,222,965
- incertae sedis 8,652,361
- Protozoa 2,511,432
- Viruses 794,566
- Archaea 373,881

TAXONOMIC DISTRIBUTION OF OCCURRENCES



SPECIES

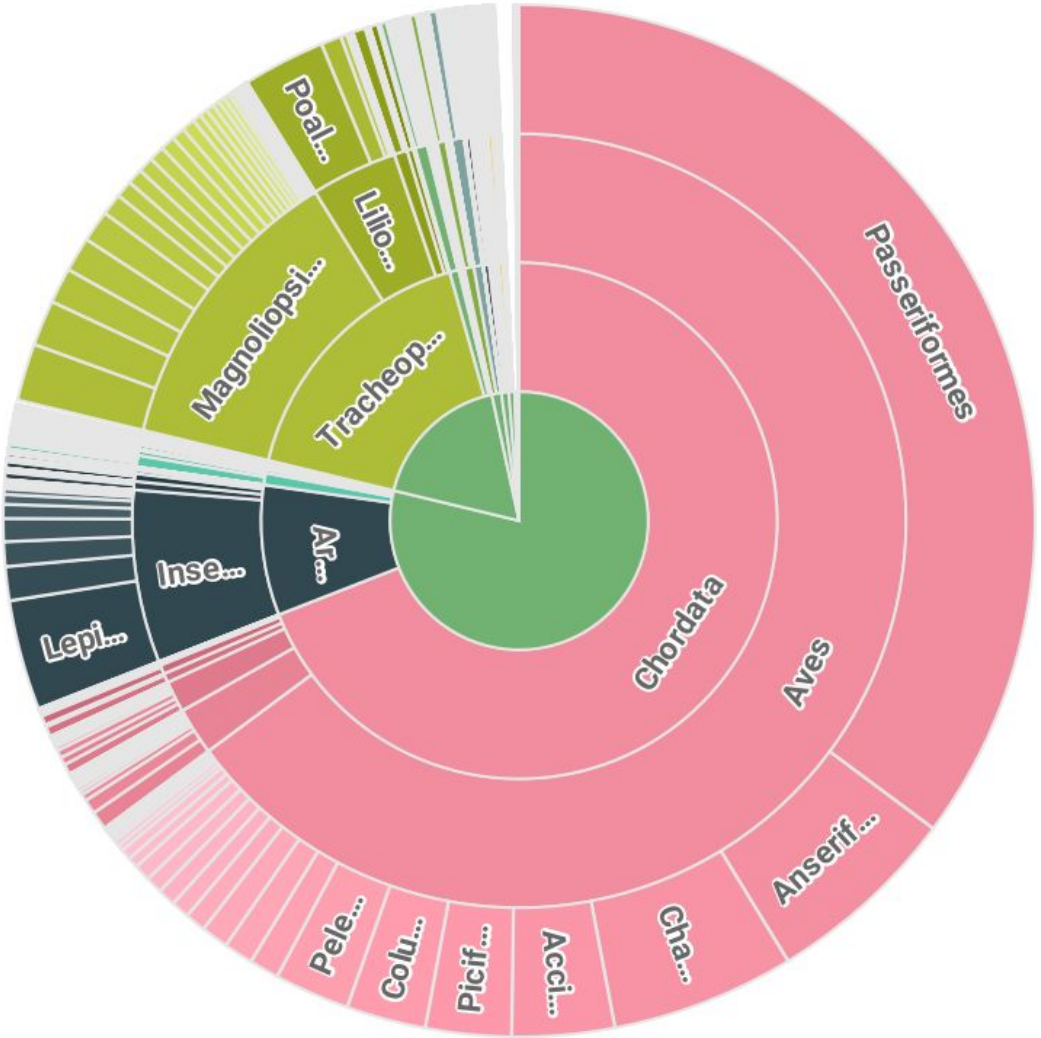
Group by Species

	Occurrences	Scientific name	Kingdom	Phylum	Class	Order	Family	Genus	Species
+	22,142,715	Anas platyrhynchos Linnaeus, 1758	Animalia	Chordata	Aves	Anseriformes	Anatidae	Anas	
+	19,546,848	Zenaida macroura (Linnaeus, 1758)	Animalia	Chordata	Aves	Columbiformes	Columbidae	Zenaida	
+	18,292,169	Corvus brachyrhynchos C.L.Brehm, 1822	Animalia	Chordata	Aves	Passeriformes	Corvidae	Corvus	
+	18,225,608	Turdus migratorius Linnaeus, 1766	Animalia	Chordata	Aves	Passeriformes	Turdidae	Turdus	

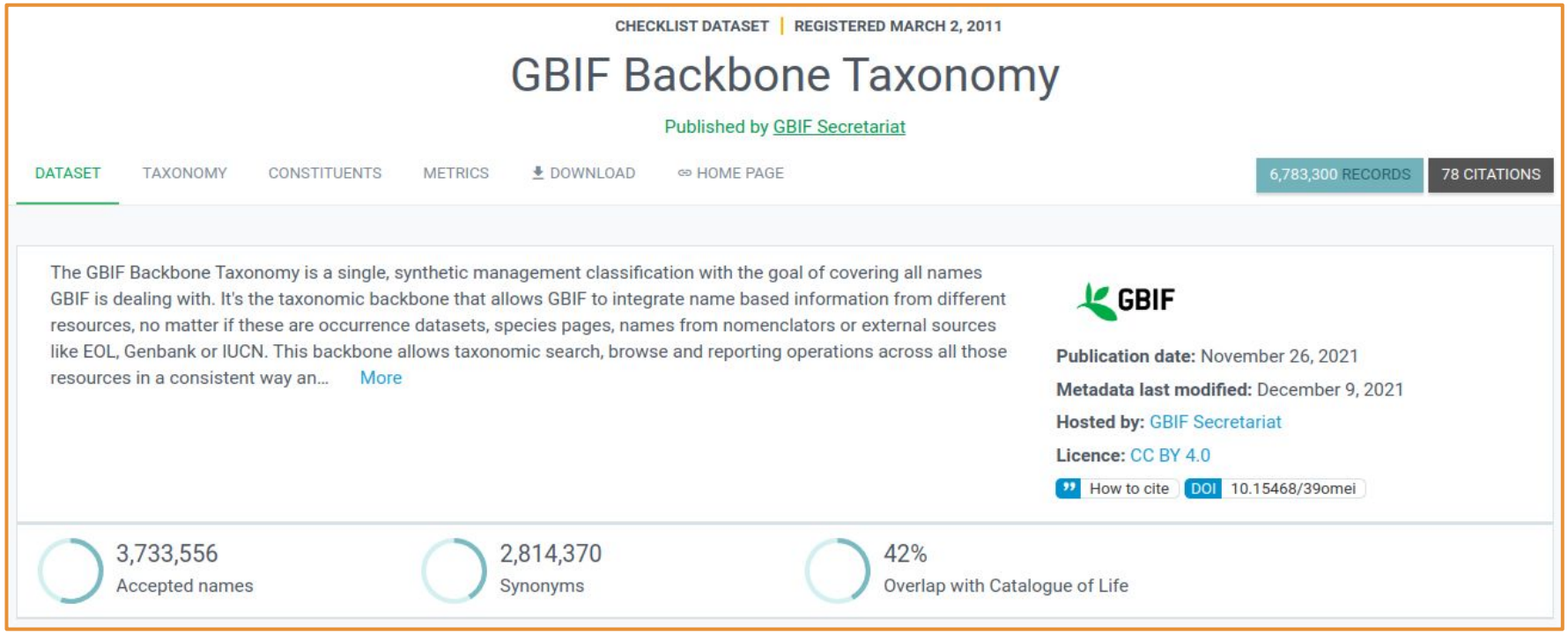
Что ?

три основные группы:

Птицы
Сосудистые растения
Насекомые



GBIF Taxonomy Backbone



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

ITIS

Integrated Taxonomic Information System

1996

SPECIES 2000

international BARCODE OF LIFE

Barcode Index Numbers (BINs) - 562985 записей

Systema Dipterorum

The BioSystematic Database of World Diptera

251 595 записей

WoRMS

World Register of Marine Species

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

Biodiversity Data Quality

Biodiversity Informatics Curriculum

Biodiversity Services and Clients

Biological Interactions Data

Citizen Science

Collection Descriptions

Darwin Core

Earth Sciences and Paleobiology

Genomic Biodiversity

Machine Observations

Observations & specimens

Process

Conferences

TDWG 2021 - Virtual Annual Conference

TDWG 2020 - A Virtual Conference

2019 - Leiden, The Netherlands

2018 - Dunedin, New Zealand

2017 - Ottawa, Canada

2016 - Santa Clara de San Carlos, Costa Rica

2015 - Nairobi, Kenya

2014 - Jönköping, Sweden

2013 - Florence, Italy

2012 - Beijing, China

2011 - New Orleans, Louisiana, USA

2010 - Woods Hole, Massachusetts, USA

Стандарт Darwin Core (DwC)

TDWG Home Terms Guides Namespace policy

Darwin Core quick reference guide

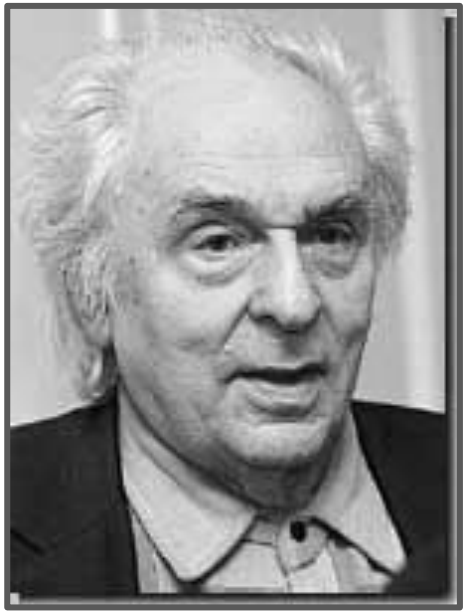
This page provides a list of all currently recommended terms of the Darwin Core standard. Categories such as `Occurrence` or `Event` correspond to Darwin Core classes which group other terms. Convenient files of these terms and their full history can be found in the Darwin Core repository.

Record-level

type | modified | language | license | rightsHolder | accessRights | bibliographicCitation | references
institutionID | collectionID | datasetID | institutionCode | collectionCode | datasetName
ownerInstitutionCode | basisOfRecord | informationWithheld | dataGeneralizations | dynamicProperties

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

ЗООЛОГИЧЕСКИЙ ИНСТИТУТ РАН
1832

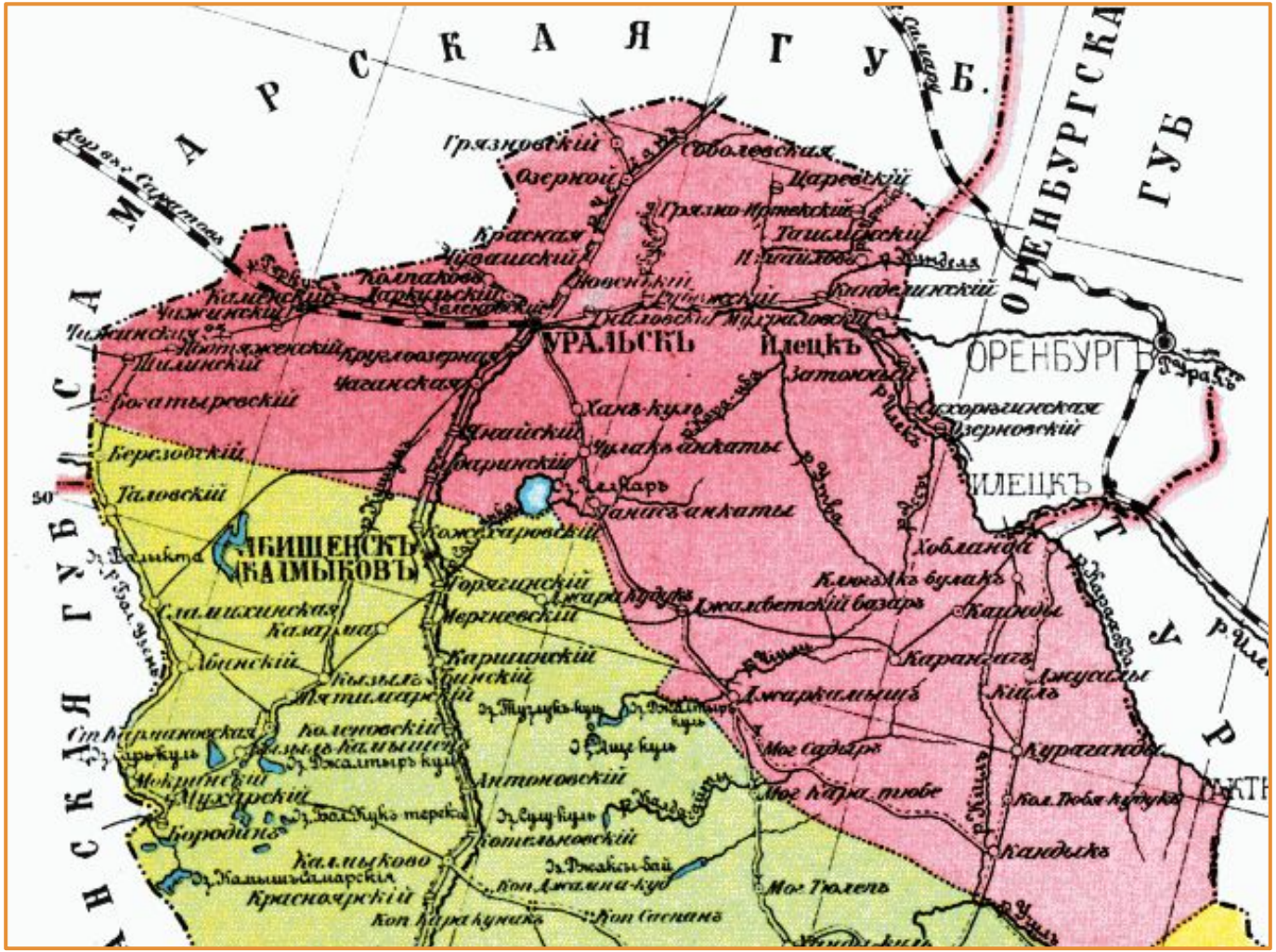
метод представления таксономических иерархических классификаций в реляционных базах данных (концепция ZOOCOD)

Где?

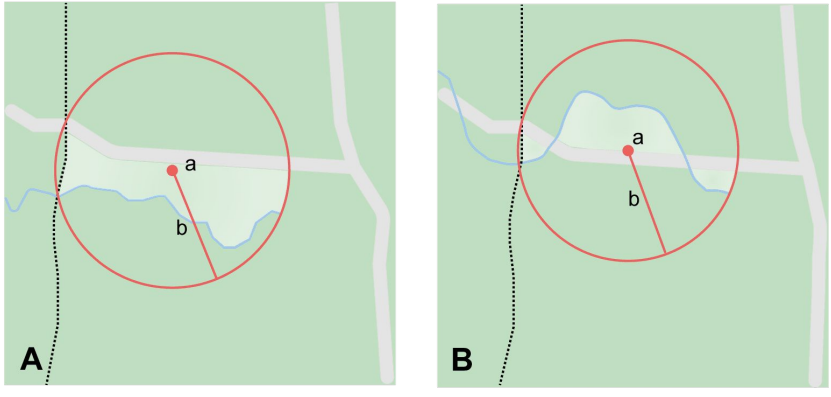
Более 90 % имеют географическую привязку

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

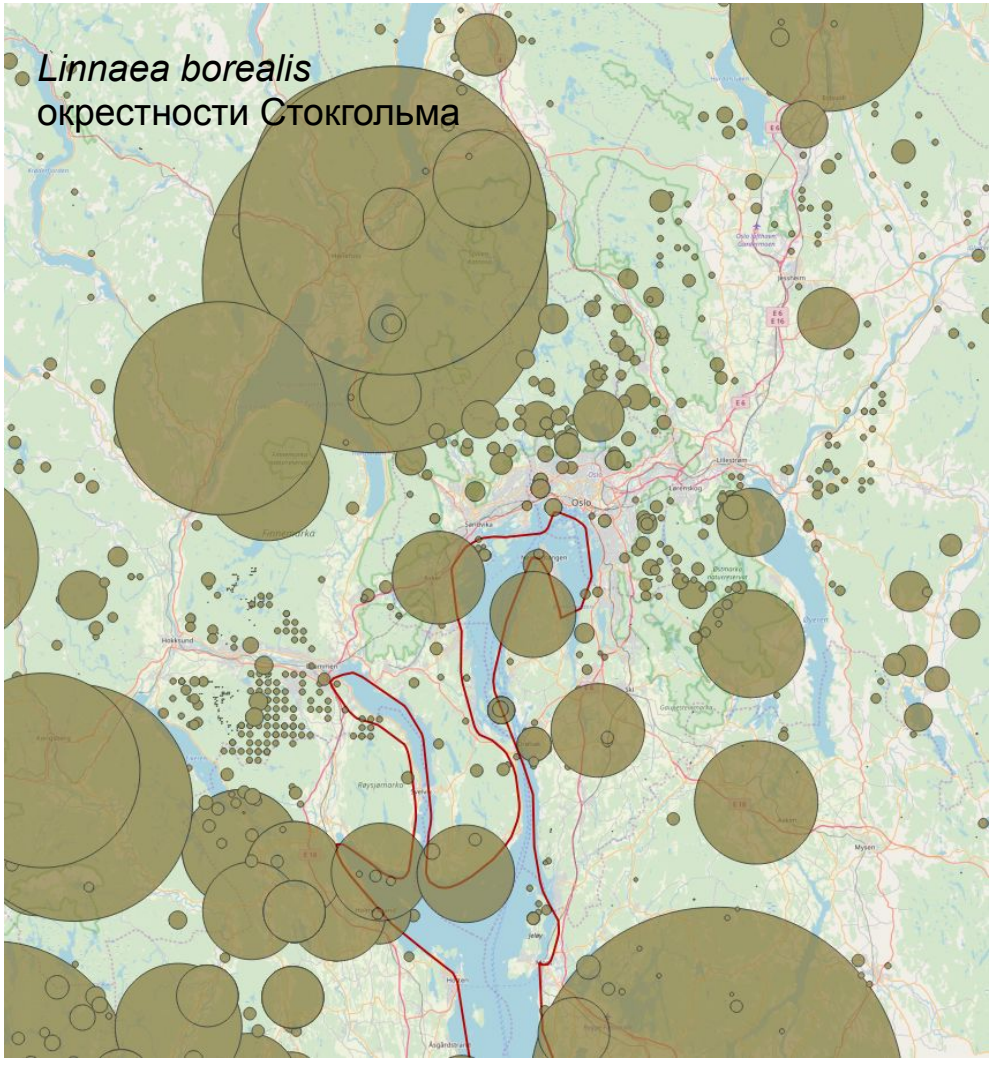
точность привязки до 100 м - менее 10%



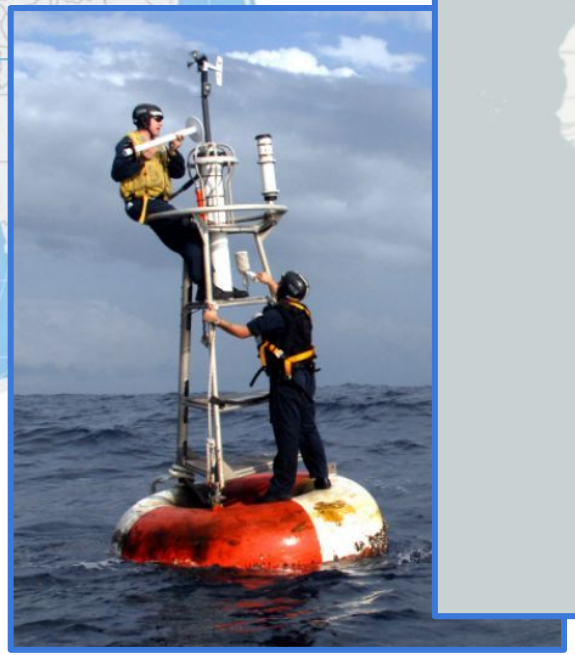
Где?
Data Quality - географическая привязка

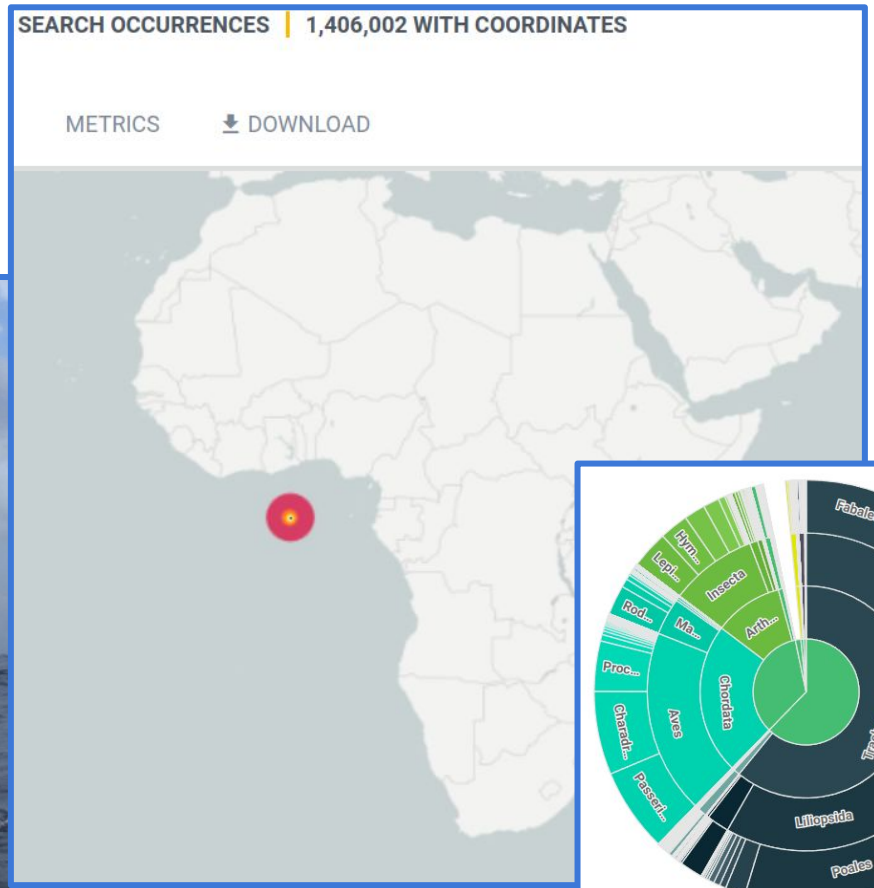


Zermoglio et al., 2020
Georeferencing Quick Reference Guide



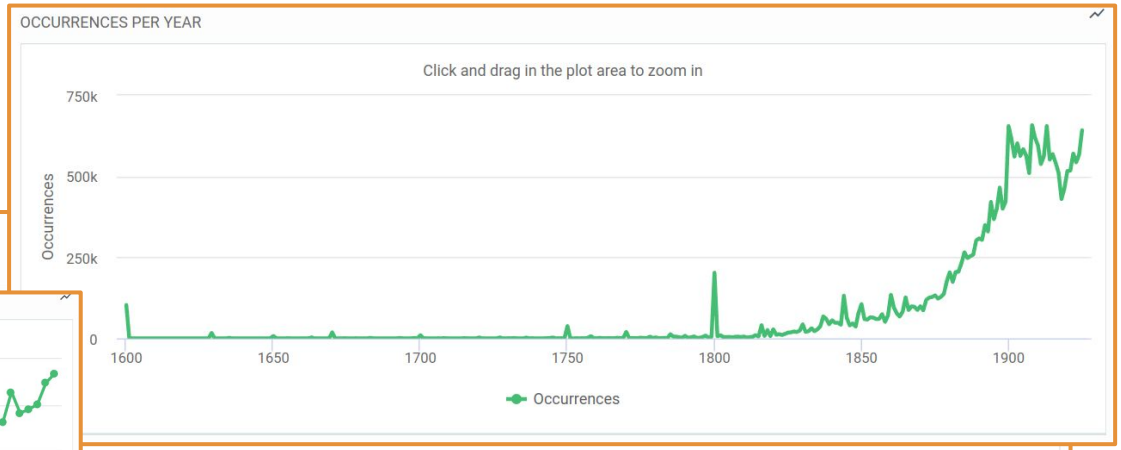
Где? Null Island POINT (0,0)

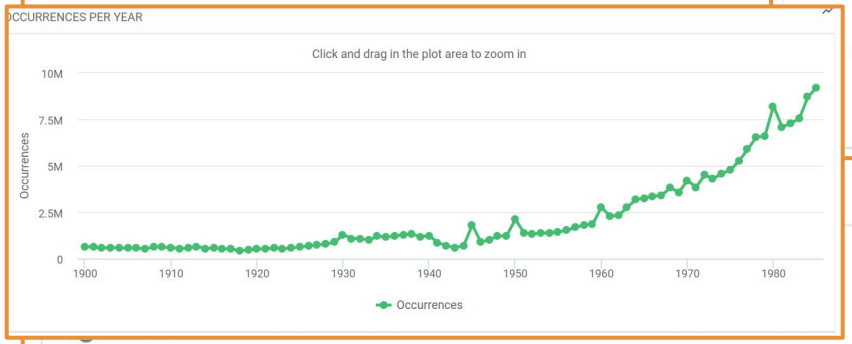




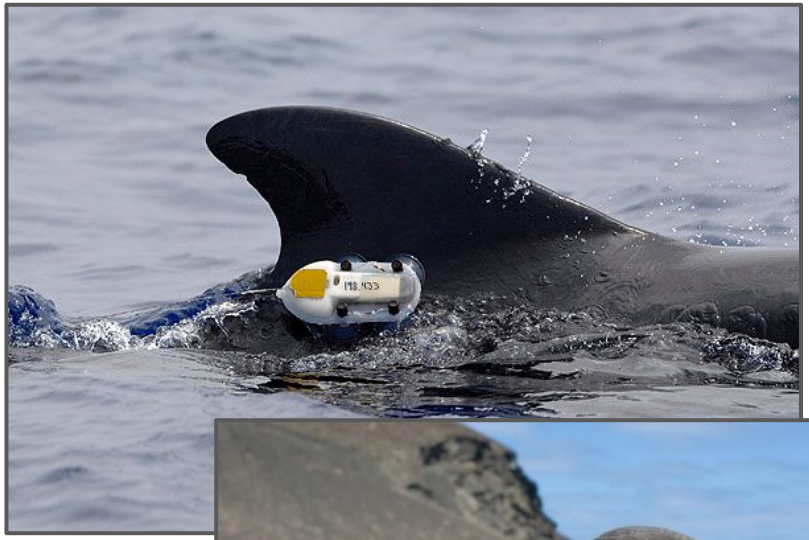
Когда?

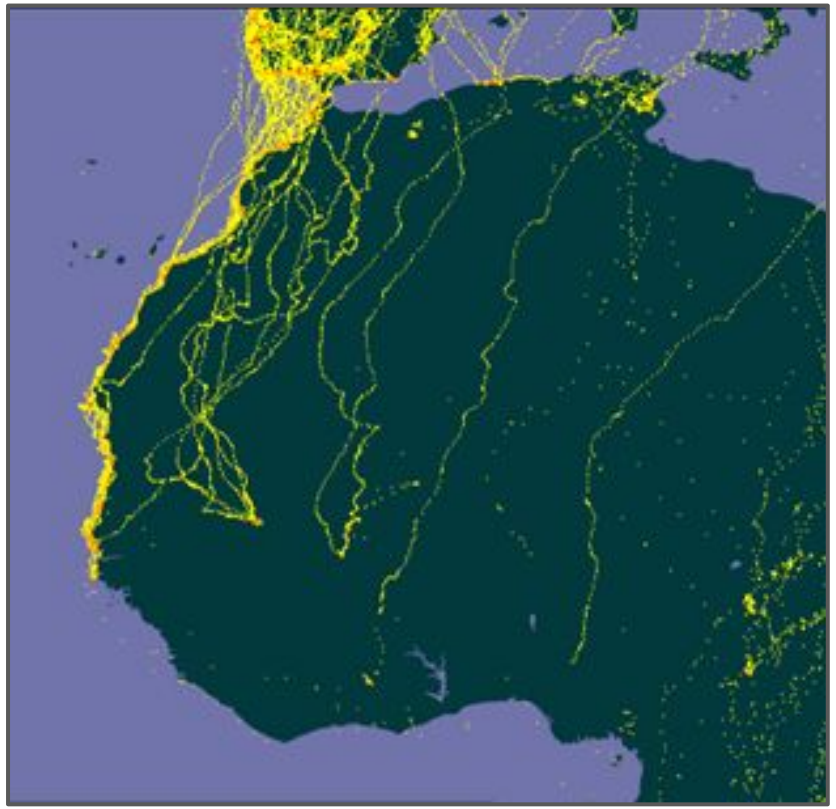
первый миллион находок
1600-1833 гг.





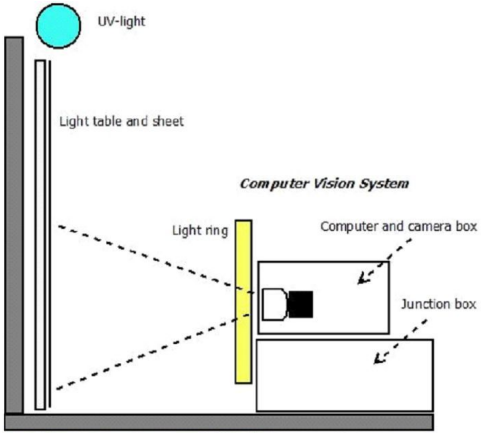
Типы данных - автоматические наблюдения

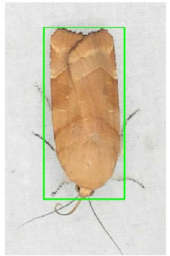


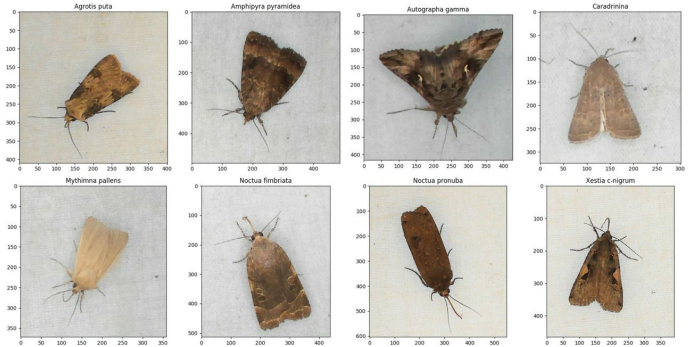


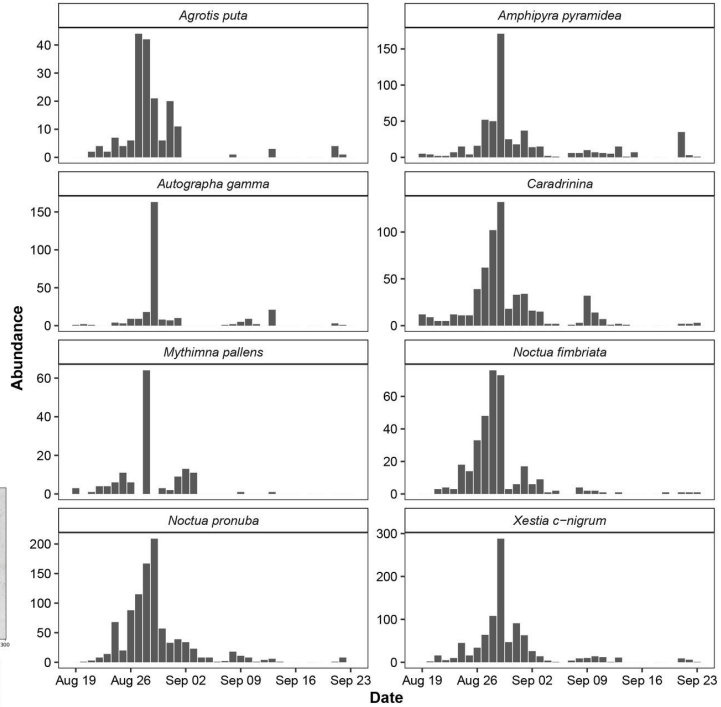
GPS-трекеры на крупных перелетных птицах

Автоматизированный учёт насекомых



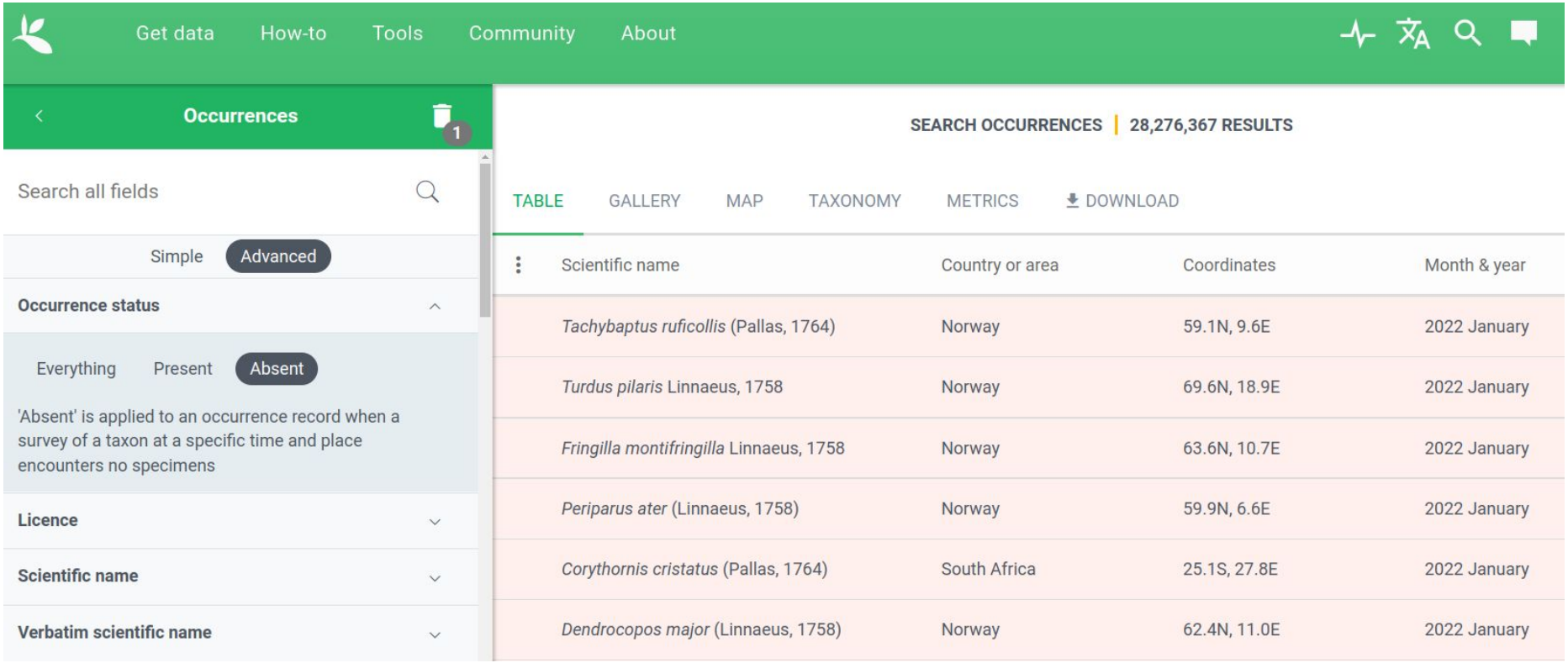






10.3390/s21020343

Точки отсутствия



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

……..

Типы наборов данных

Описание набора данных

Список видов

Отдельные находки
(количественные характеристики)

Характиристики 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

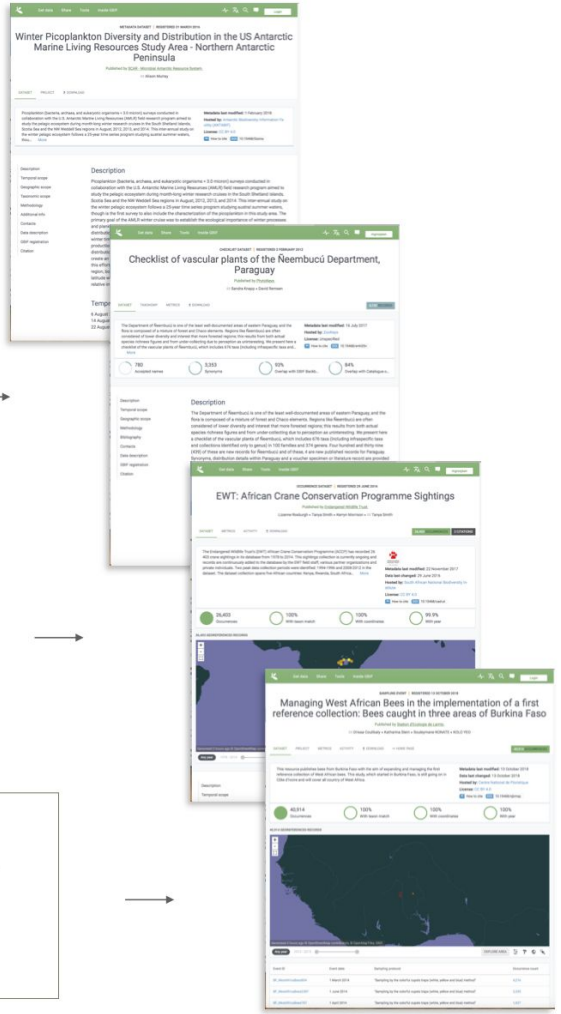
- Scientific names of organisms sharing a common theme or feature (for example: medicinal use).

Occurrence Dataset

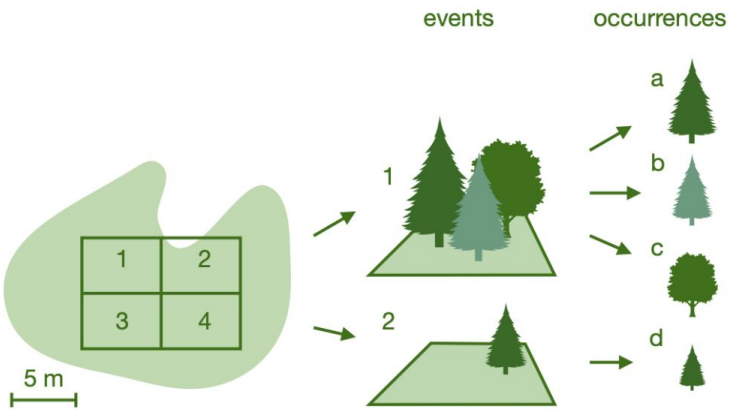
- Scientific names of organisms observed or specimens collected,
- Observation or sampling date (year),
- Observation or sampling location (at least country).

Sampling-Event Dataset

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



Sampling Event dataset



Occurrence file

occurrenceID	scientificName	EventID	basisOfRecord	…
a	species A	1	humanObservation	
b	species B	1	humanObservation	
c	species C	1	humanObservation	
d	species A	2	humanObservation	
…				

Event file

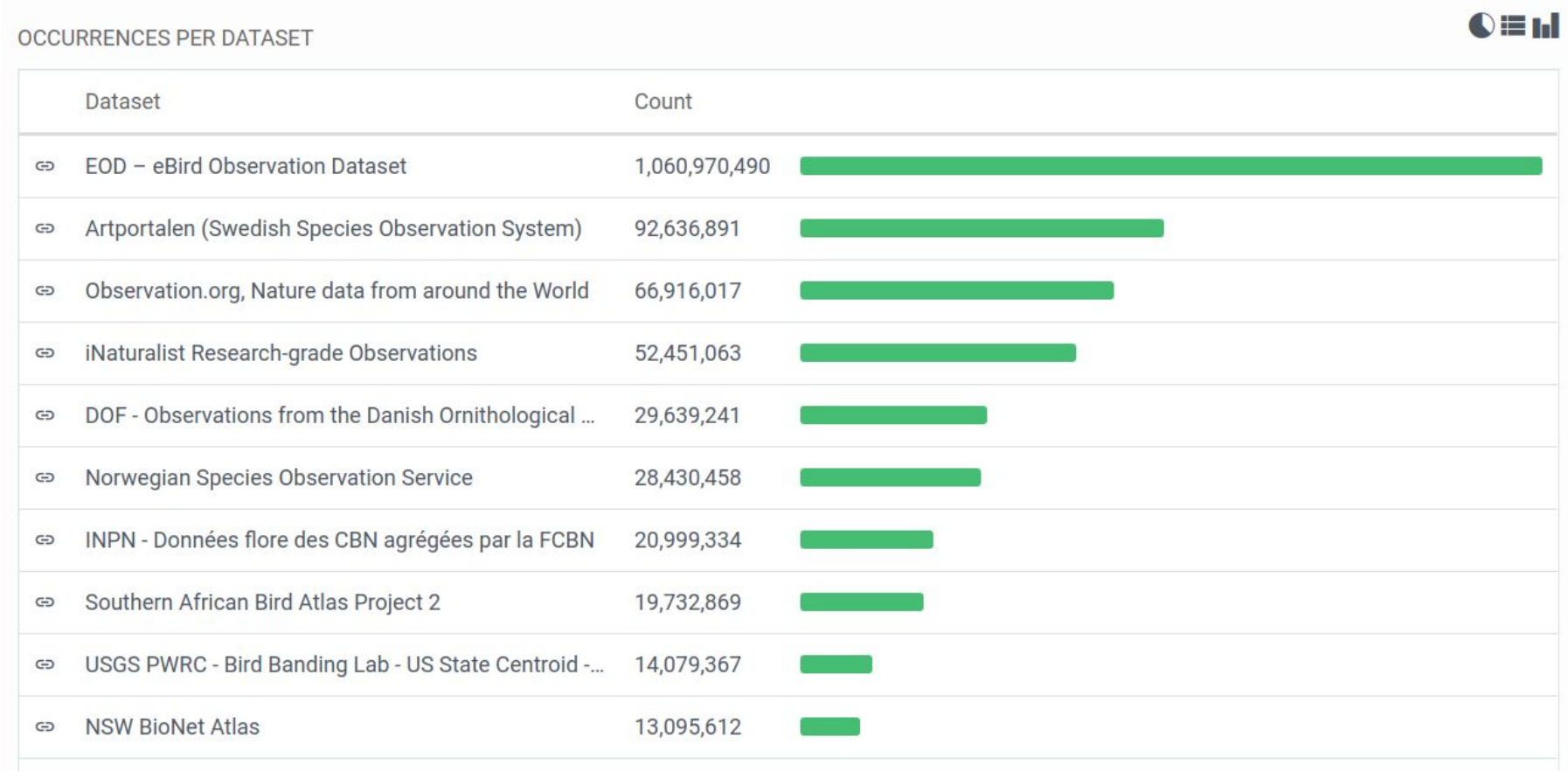
EventID	EventDate	samplingProtocol	sampleSizeValue	sampleSizeUnit	…
1	2018-08-03	ad hoc observations	25	square metre	
2	2018-08-03	ad hoc observations	25	square metre	
…					

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

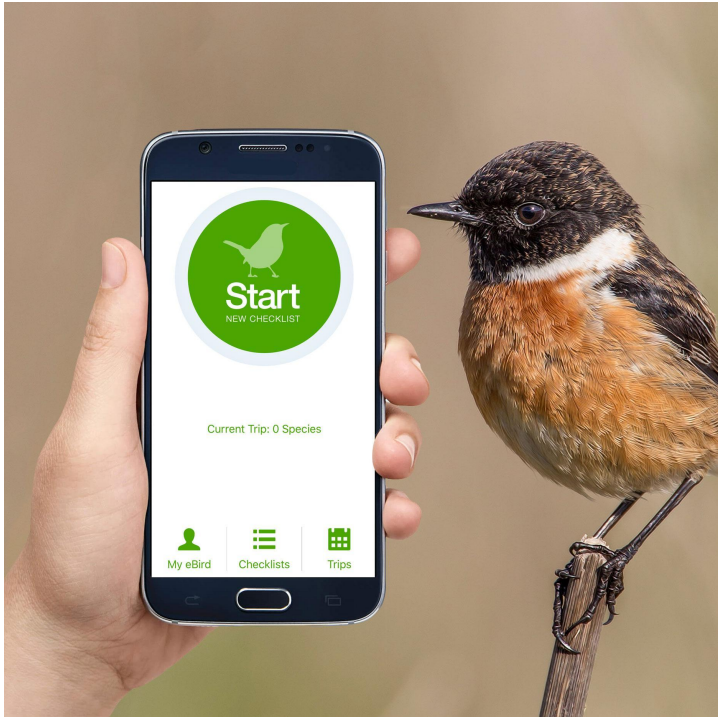
Источники данных



Всемирная сеть наблюдений за птицами

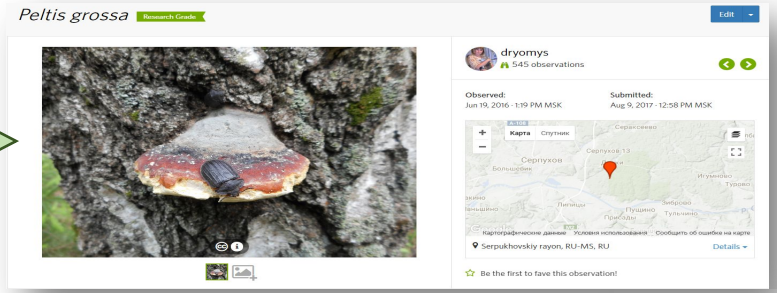
> 1 000 000 000 находок

https://ebird.org/home

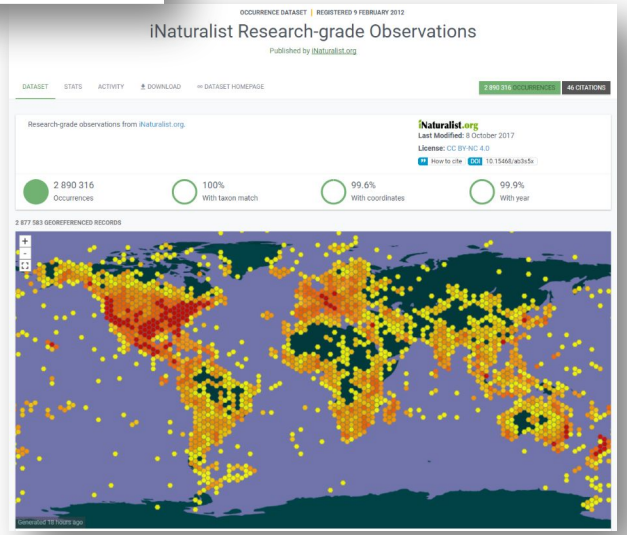


Сообщество натуралистов – система Inaturalist.org

~ 74 млн Research Grade наблюдений



> 52 млн. записей



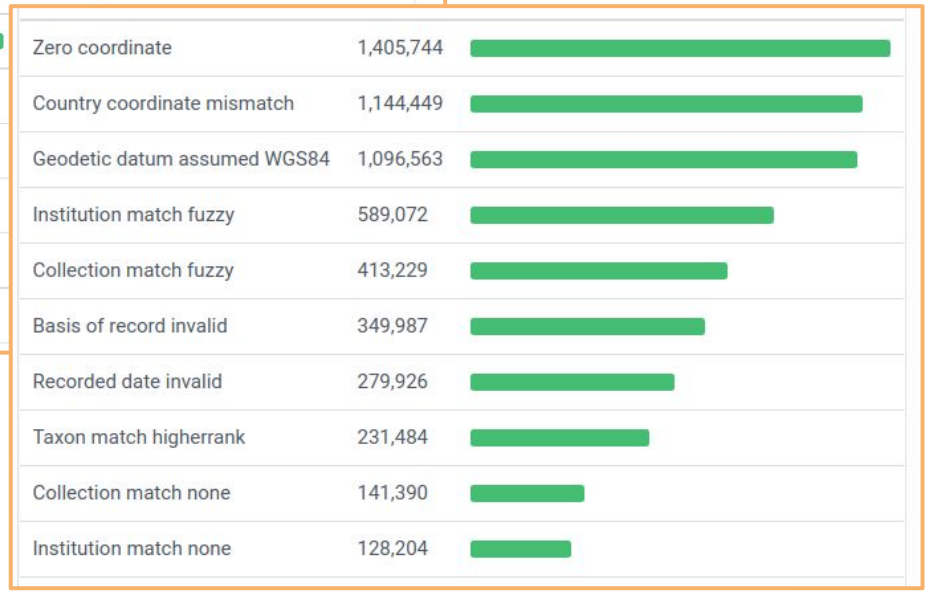
iNaturalist.org

~ 2.5 млн. участников

> 120 млн наблюдений

GBIF.org

Качество данных - Data Quality



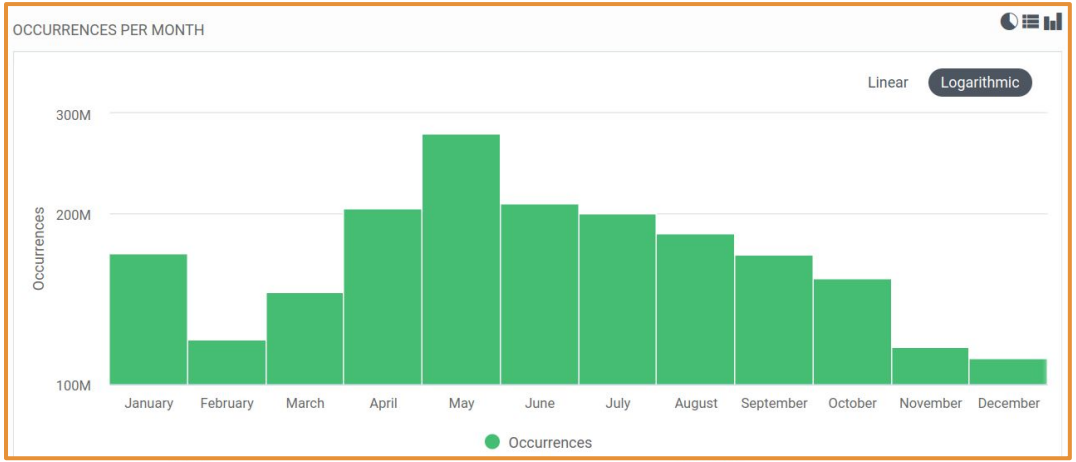
Качество данных - Data Quality

Occurrence-only datasets

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

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

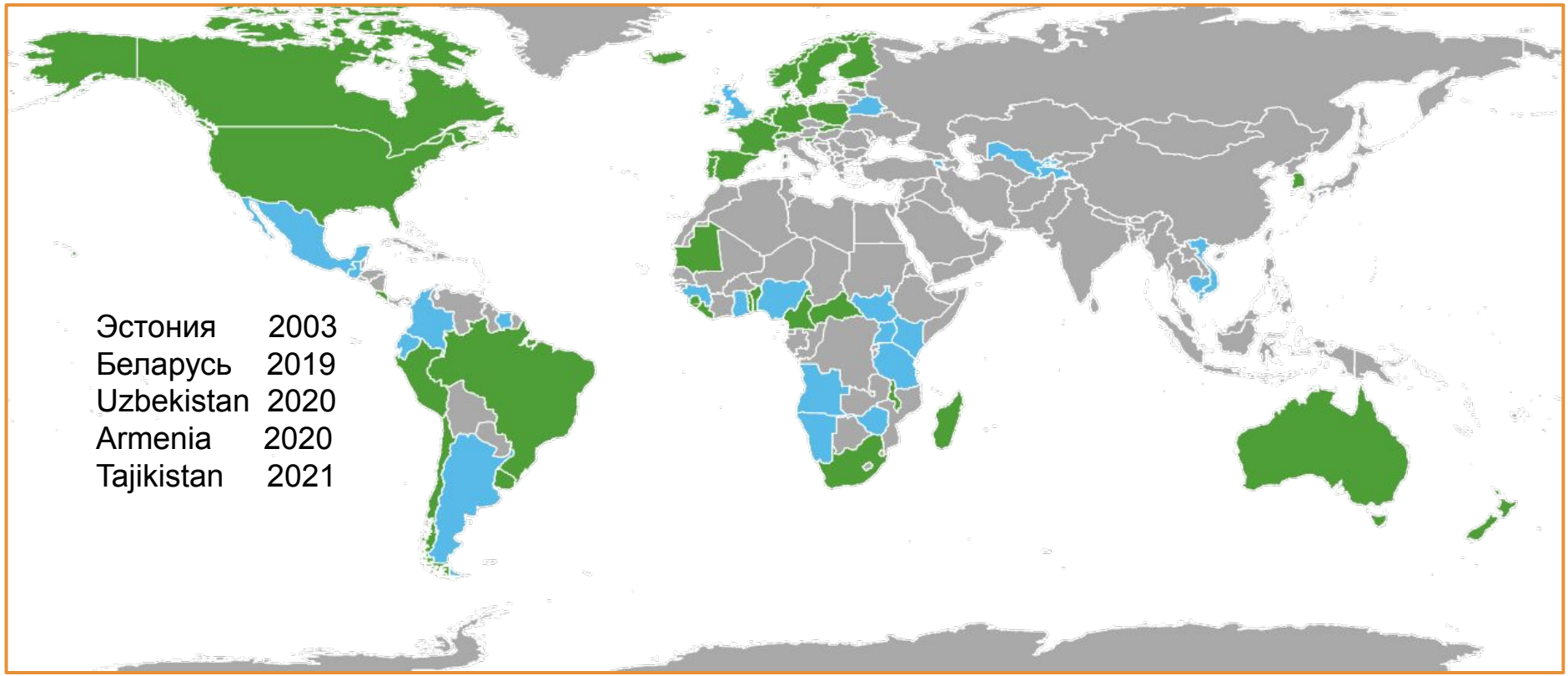


Checklist datasets

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

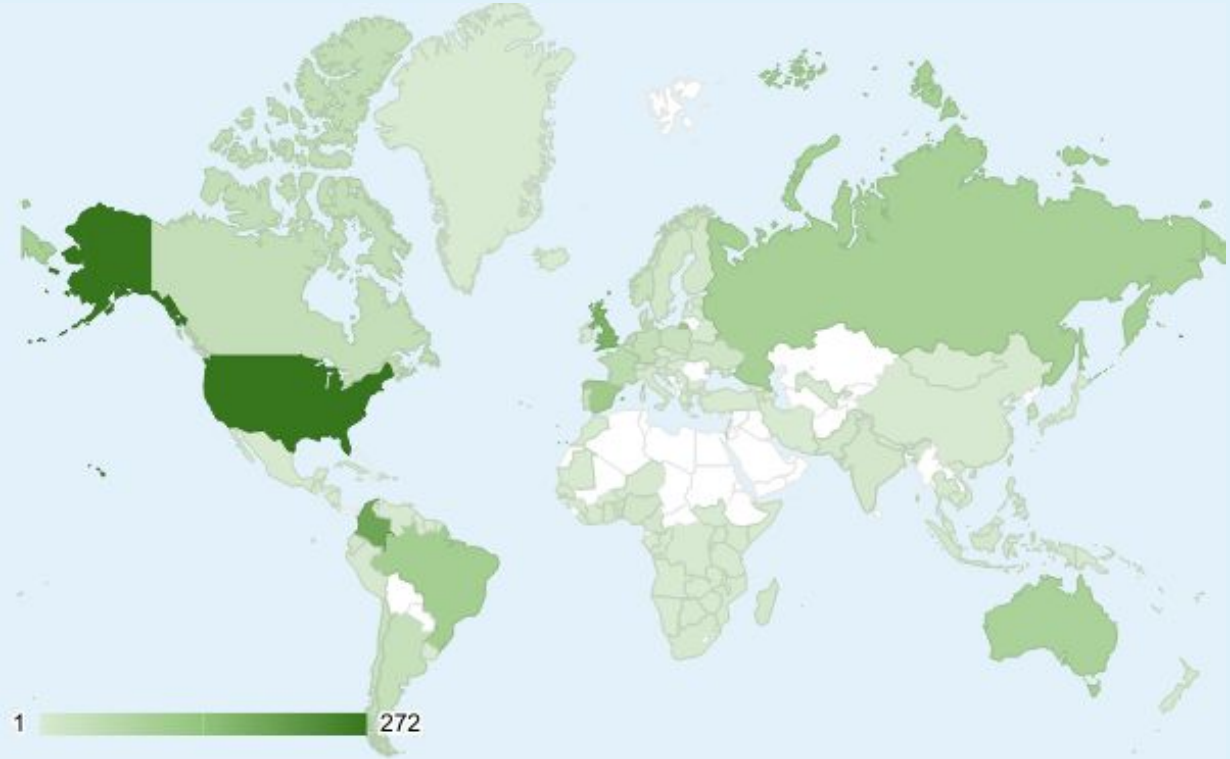
GBIF - не только портал



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

https://www.gbif.org/publisher/search

GBIF

Российские участники GBIF

на начало 2019 года





















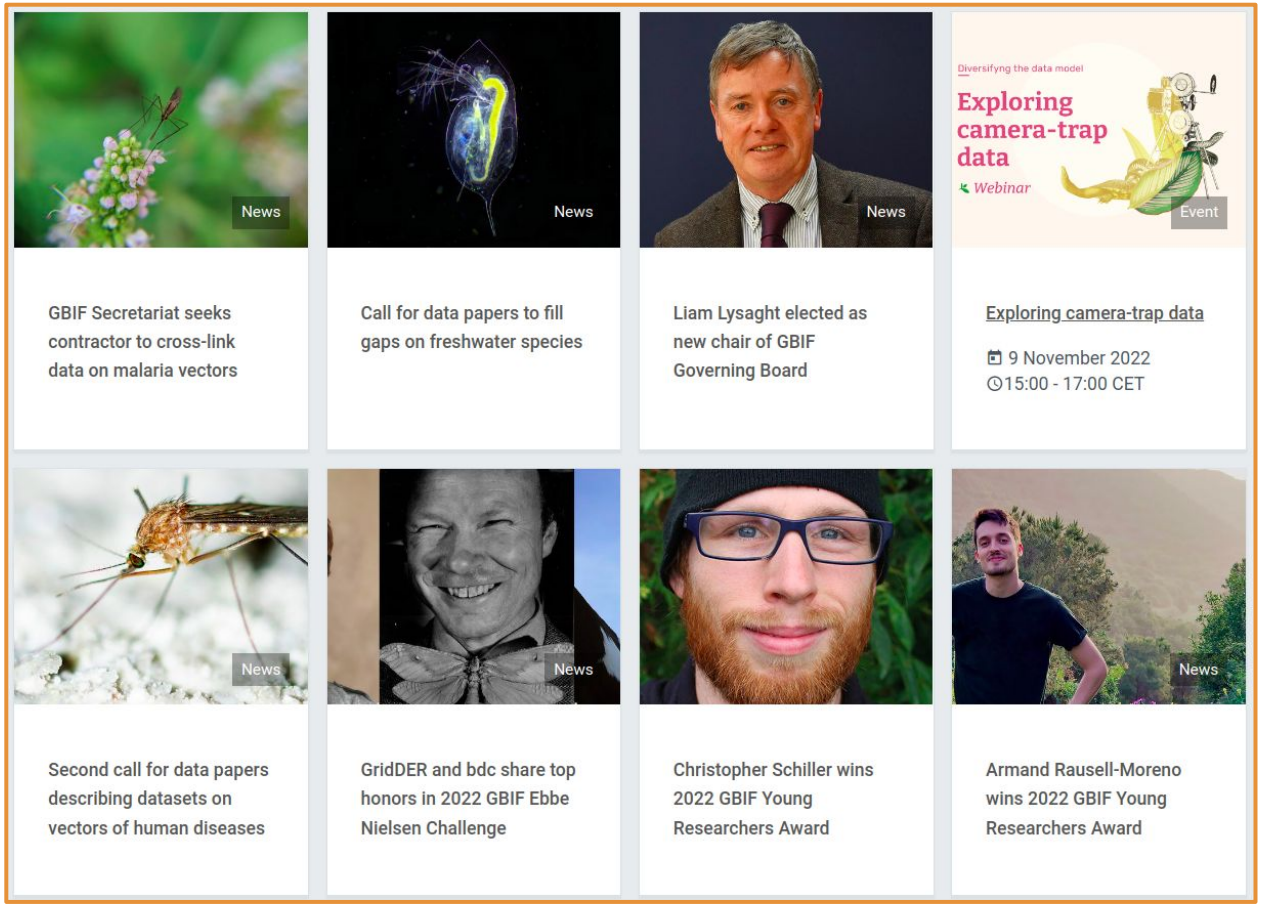
Новости портала GBIF

Вебинары

Учебные курсы

Конкурсы поддержки проектов

Премии для молодых учёных



Data papers: Bringing data to light — 8 ноября

9 ноября

Exploring camera-trap data