



# Биоэртүрлілікті зерттеудегі цифрлық технологиялар

## Digital technologies in biodiversity research

### Цифровые технологии в исследовании биоразнообразия



#### ЛЕКЦИЯ 1

## Полевая биология в цифровом мире



Слайды CC BY:

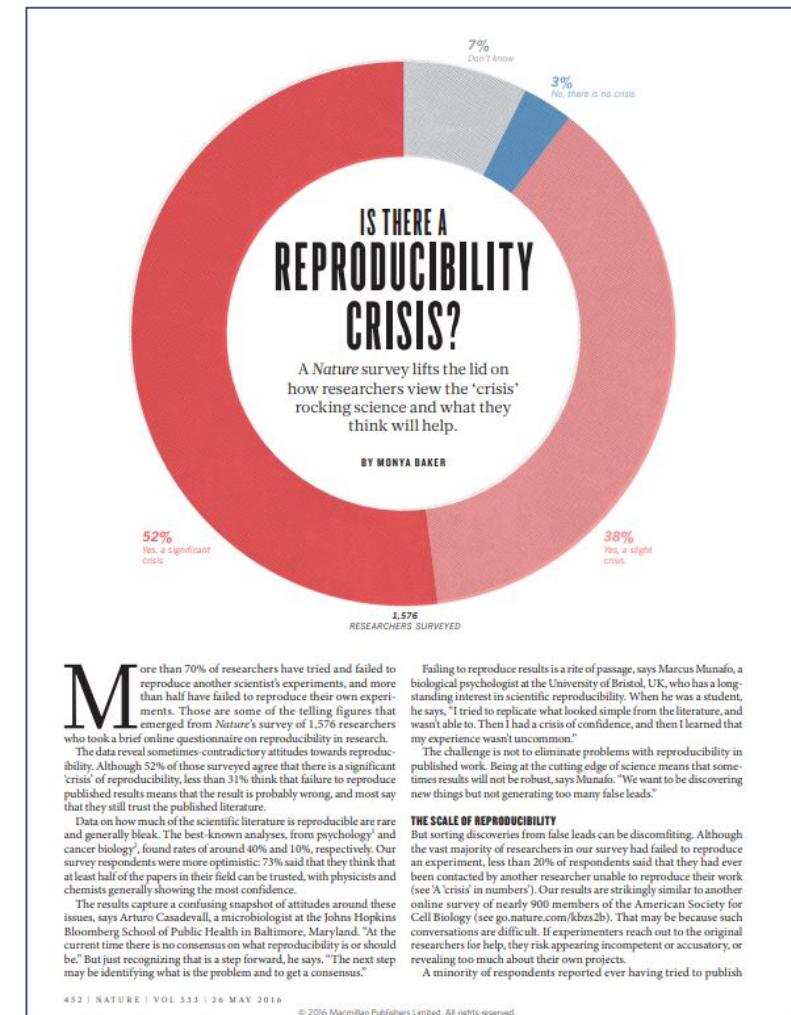
Dag Endresen, GBIF Norway  
Артём Созонтов, ИЭРИЖ УрО РАН  
Наталья Иванова,  
Максим Шашков



Natalya Ivanova

# Есть ли кризис воспроизводимости научных исследований?

Результаты опроса 1576 исследователей показали, что более 70% исследователей пытались и не смогли воспроизвести эксперименты другого ученого, и более половины не смогли воспроизвести свои собственные эксперименты.



**M**ore than 70% of researchers have tried and failed to reproduce another scientist's experiments, and more than half have failed to reproduce their own experiments. Those are some of the telling figures that emerged from *Nature*'s survey of 1,576 researchers who took a brief online questionnaire on reproducibility in research.

The data reveal sometimes-contradictory attitudes towards reproducibility. Although 52% of those surveyed agree that there is a significant 'crisis' of reproducibility, less than 31% think that failure to reproduce published results means that the result is probably wrong, and most say that they still trust the published literature.

Data on how much of the scientific literature is reproducible are rare and generally bleak. The best-known analyses, from psychology<sup>1</sup> and biology<sup>2</sup>, found rates of around 40% and 10%, respectively. Our survey respondents were more optimistic: 73% said that they think that at least half of the papers in their field can be trusted, with physicists and chemists generally showing the most confidence.

The results capture a confusing snapshot of attitudes around these issues, says Arturo Casadevall, a microbiologist at the Johns Hopkins Bloomberg School of Public Health in Baltimore, Maryland. "At the current time there is no consensus on what reproducibility is or should be." But just recognizing that is a step forward, he says. "The next step may be identifying what is the problem and to get a consensus."

A minority of respondents reported ever having tried to publish

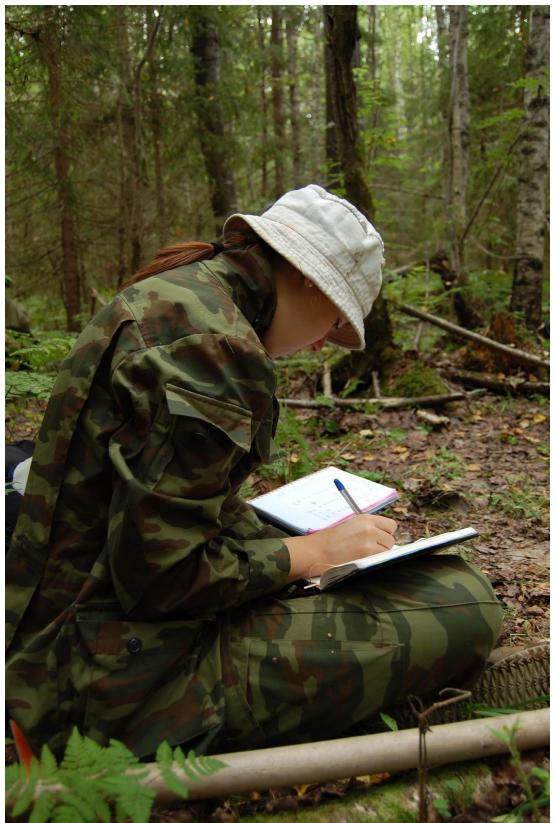
Failing to reproduce results is a rite of passage, says Marcus Munafò, a biological psychologist at the University of Bristol, UK, who has a long-standing interest in scientific reproducibility. When he was a student, he says, "I tried to replicate what looked simple from the literature, and wasn't able to. Then I had a crisis of confidence, and then I learned that my experience wasn't uncommon."

The challenge is not to eliminate problems with reproducibility in published work. Being at the cutting edge of science means that sometimes results will not be robust, says Munafò. "We need to be discovering new things but not generating too many false leads."

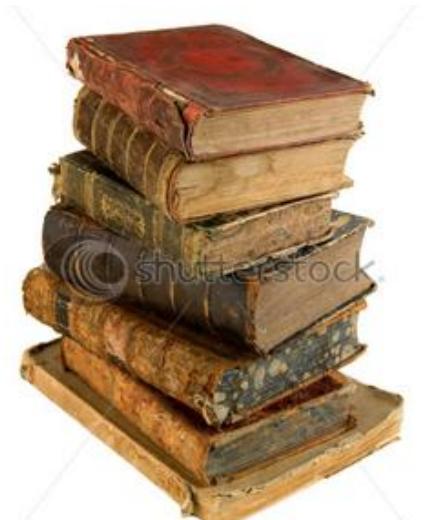
**THE SCALE OF REPRODUCIBILITY**  
But sorting discoveries from false leads can be disconcerting. Although the vast majority of researchers in our survey had failed to reproduce an experiment, less than 20% of respondents said that they had ever been contacted by another researcher unable to reproduce their work (see 'A crisis in numbers'). Our results are strikingly similar to another online survey of nearly 900 members of the American Society for Cell Biology (see go.nature.com/khsz2b). That may be because such conversations are difficult. If experimenters reach out to the original researchers for help, they risk appearing incompetent or accusatory, or revealing too much about their own projects.

A minority of respondents reported ever having tried to publish

Baker, M. 1,500 scientists lift the lid on reproducibility. *Nature* 533, 452–454 (2016). <https://doi.org/10.1038/533452a>

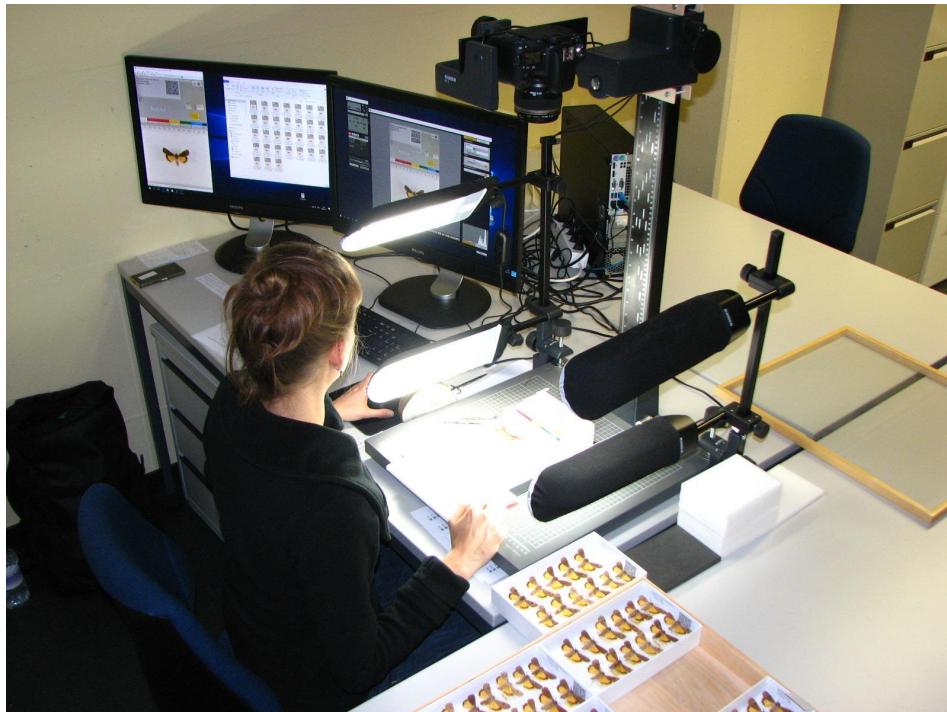


Полевые данные ценные сами по себе и должны быть  
доступны для повторного анализа



# Оцифровка научных биологических коллекций

Making data and images of millions of biological specimens available on the web



# Biodiversity Heritage Library

<https://www.biodiversitylibrary.org/>



- Открытый доступ к литературе о биоразнообразии
- >147000 источников, опубликованных с 1450 по 2022 гг.

Biodiversity  
Heritage  
Library



*Inspiring discovery through free access to biodiversity knowledge.*

The Biodiversity Heritage Library improves research methodology by collaboratively making biodiversity literature openly available to the world as part of a global biodiversity community.

BHL also serves as the foundational literature component of the Encyclopedia of Life (EOL).

Search across books and journals, scientific names, authors and subjects

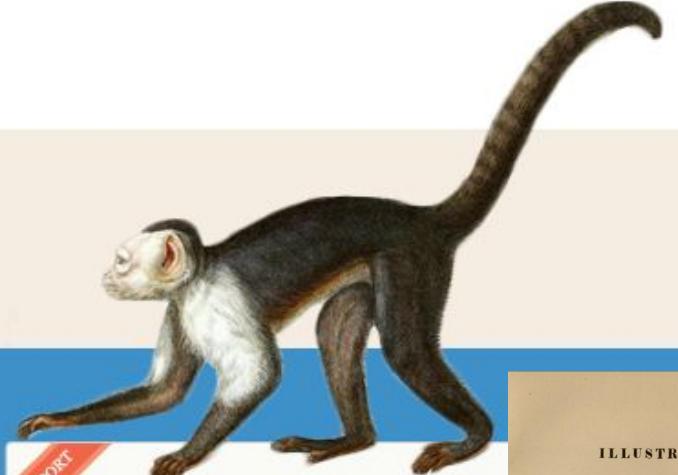
Search the catalog and full-text 

Full-text  Catalog [?](#)

[ADVANCED SEARCH](#)

Browse by:

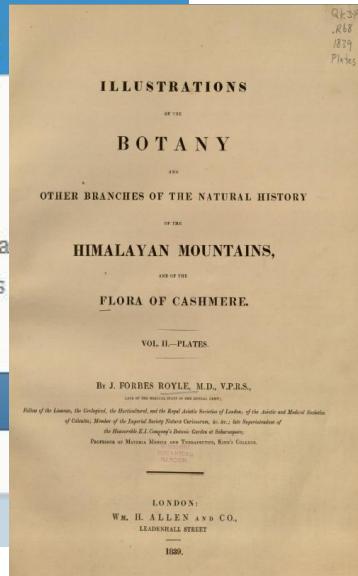
[Title](#) [Author](#) [Date](#) [Collection](#) [Contributor](#)



**Help Support BHL**

BHL's existence depends on the financial support of its patrons. Help us keep this free resource alive!

[Donate Now](#)





# Citizen science

Привлечение волонтеров к сбору научных  
данных под руководством  
профессиональных исследователей

Подробнее: [https://en.wikipedia.org/wiki/Citizen\\_science](https://en.wikipedia.org/wiki/Citizen_science)

Объем данных, собираемых через системы  
любительских наблюдений,  
в мире возрастает существенно быстрее по  
сравнению с «научными» источниками

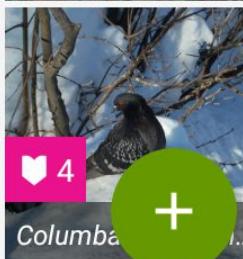


Для данных, доступных через GBIF

- EOD – eBird Observation Dataset  
> 1 млрд наблюдений
- Observation.org > 58 млн наблюдений
- iNaturalist Research-grade Observations  
> 52 наблюдений

My Observati... ≡ 🔍 ⋮

4,761 OBSERVATIONS 1,027 SPECIES 1,431 IDENTIFICATIONS

 <i>Catocala fraxini</i>	 <i>Plantago media</i>	 <i>Gagea minima</i>
 <i>Ficaria verna</i>	 <i>Corydalis solida</i>	 <i>Petasites spurius</i>
 <i>Catocala nupta</i>	 <i>Bromus inermis</i>	 <i>Chelidonium majus</i>
 <i>Vicia sepium</i>	 <i>Trifolium pratense</i>	 <i>Columba...</i>

# Мобильное приложение iNaturalist

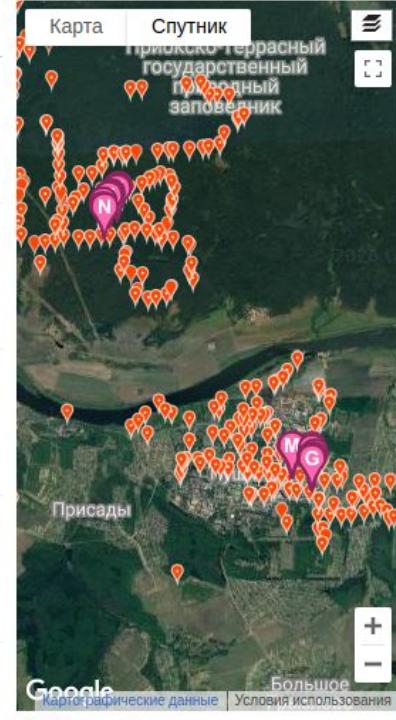
Explore Your Observations Community Identify More

Your Observations

Add Observations Batch edit Search

Photos / Sounds	Species / Taxon Name	Date observed	Place	Date Added
 A	<i>Taraxacum officinale</i> Common Dandelion	May 16, 2020 10:04 AM MSK	8, Pushchino, Moskovskaya oblast', Russia, 142290 (Google, OSM)	May 16, 2020 10:53 AM MSK 1 ID Research Grade Edit   View »
 B	<i>Taraxacum officinale</i> Common Dandelion	May 16, 2020 09:54 AM MSK	Mkrn. D., Pushchino, Moscow Oblast, Russia, 142290 (Google, OSM)	May 16, 2020 10:52 AM MSK 1 ID Research Grade Edit   View »
 C	<i>Artemisia vulgaris</i> Common Mugwort	May 16, 2020 09:52 AM MSK	Mkrn. D., Pushchino, Moscow Oblast, Russia, 142290 (Google, OSM)	May 16, 2020 10:47 AM MSK 1 ID Research Grade Edit   View »
 D	<i>Chelidonium majus</i> Greater Celandine	May 16, 2020 09:51 AM MSK	Mkrn. D., Pushchino, Moscow Oblast, Russia, 142290 (Google, OSM)	May 16, 2020 10:47 AM MSK 1 ID Research Grade Edit   View »
 E	<i>Lamium album</i> White Deadnettle	May 16, 2020 09:51 AM MSK	Mkrn. D., Pushchino, Moscow Oblast, Russia, 142290 (Google, OSM)	May 16, 2020 10:46 AM MSK 2 IDs Research Grade Edit   View »

Карта Спутник



Google Картографические данные Условия использования

Redo search in map area

Версия, доступная через браузер [www.inaturalist.org/home](http://www.inaturalist.org/home)

# FAIR принципы данных

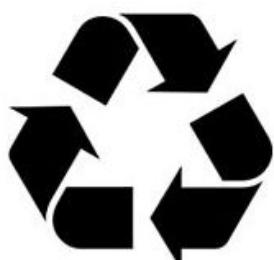
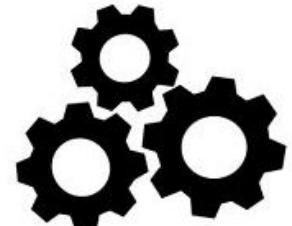
Данные можно найти в сети Интернет  
! идентификаторы !  
! метаданные !

Findable  
A  
ccessible



Данные должны быть совместимыми с другими подобными массивами данных  
! стандарты !

Interoperable  
R  
Reusable



Данные должны быть доступны  
! бесплатно и в открытом доступе !

Данные могут быть многократно использованы для решения разных научных и практических задач  
! лицензии (правила использования) !

# Политика научных журналов в отношении исходных данных: обзор издательства

**SPRINGER NATURE**

## Тип 1

Журнал призывает авторов, когда это возможно и целесообразно, размещать данные, подтверждающие результаты их исследований, в общедоступном хранилище.

## Тип 2

Журнал настоятельно рекомендует, чтобы все наборы данных, на которые опираются выводы статьи, были доступны для читателей. Авторам рекомендуется предоставить информацию о доступности данных в своей статье.

## Тип 3

Журнал настоятельно рекомендует, чтобы все наборы данных, на которые опираются выводы статьи, были доступны для читателей. Авторам необходимо предоставить информацию о доступности данных в своей статье.

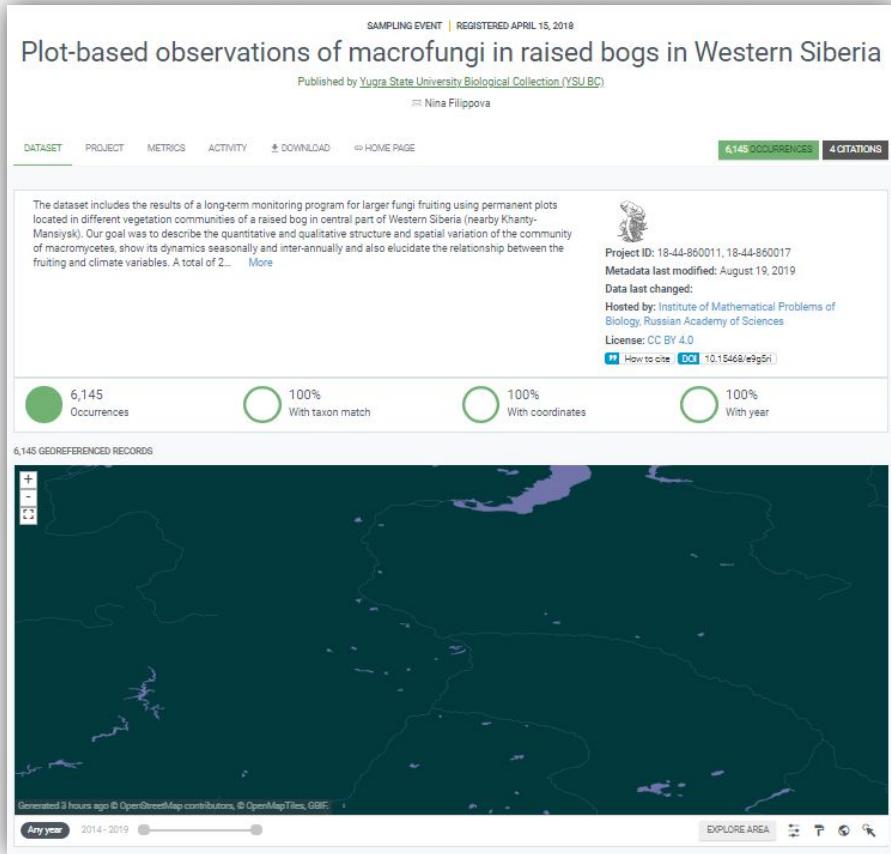
## Тип 4

Журнал требует, чтобы все наборы данных, на которые опираются выводы статьи, были доступны рецензентам и читателям. Авторам необходимо предоставить информацию о доступности данных в своей статье.

Файлы на вашем рабочем компьютере или USB-накопителе возможно содержат ценные знания, которые могут извлечь другие исследователи. <https://www.springernature.com/gp/authors/research-data>

# Data paper Статья о данных

Рецензируемая публикация в научном журнале, содержащая описание набора данных, доступного через Интернет



Biodiversity Data Journal 7: e35674  
doi: 10.3897/BDJ.7.e35674

open access

Data Paper

**Sampling event dataset on five-year observations of macrofungi fruit bodies in raised bogs, Western Siberia, Russia**

Nina Filippova<sup>1</sup>, Elena Lapshina<sup>1</sup>  
<sup>1</sup>Yugra State University, Khanty-Mansiysk, Russia

Corresponding author: Nina Filippova ([nfilippova@yusu.edu.ru](mailto:nfilippova@yusu.edu.ru))  
Academic editor: Dmitry Schigel  
Received: 23 Apr 2019 | Accepted: 23 Jul 2019 | Published: 30 Jul 2019  
Citation: Filippova N, Lapshina E (2019) Sampling event dataset on five-year observations of macrofungi fruit bodies in raised bogs, Western Siberia, Russia. Biodiversity Data Journal 7: e35674. <https://doi.org/10.3897/BDJ.7.e35674>

**Abstract**

**Background**

The data paper includes the results of a long-term monitoring programme for macrofungi fruiting using permanent plots located in different vegetation communities of a raised bog in central part of Western Siberia (nearby Khanty-Mansiysk). The goal was to describe the quantitative and qualitative structure and spatial variation of the community of macromycetes, show its dynamics seasonally and inter-annually and also elucidate the relationship between the fruiting and climate variables. A total of 263 circular 5 m<sup>2</sup> subplots (for a total area of 1,315 m<sup>2</sup>) were inspected weekly during vegetation seasons 2014–2018 and carpophores of different fungal taxa were counted. The resulting sampling-event dataset includes 16,569 of plot-based observations (= sampling events) with corresponding 6,011 occurrence records of macromycetes identified to species or genus level. In total, 69 species were revealed during the study. About 80% of plot-based observations contain zero records and mark absence of visible fruiting bodies in a certain plot and time.

© Filippova N, Lapshina E. This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Filippova N., Lapshina E. (2019)  
doi: 10.3897/BDJ.7.e35674

	<b>Исследовательская статья Research paper</b>	<b>Статья о данных Data paper</b>
<b>ЦЕЛЬ</b>	Проверка научной гипотезы	Описание первичных полевых данных, приведенных к требуемому формату
<b>РАЗДЕЛЫ</b>	Введение Материалы и методы Результаты Обсуждение Заключение	Введение Материалы и методы Описание данных
<b>РЕЦЕНЗИРОВАНИЕ</b>	Рецензирование текста рукописи	Аудит данных Рецензирование текста рукописи



NEWS | 23 MAY 2022

## Call for data papers describing datasets from Northern Eurasia (extended)

*GBIF partners with FinBIF and Pensoft to support publication of new datasets about biodiversity from across Northern Eurasia*



Northern banded newt (*Ommatotriton ophryticus*), observed in Georgia. Photo 2021 Natalia Bulbulashvili via iNaturalist Research-grade Observations, licensed under CC BY-NC 4.0.

- **Подача рукописей до 1 декабря**
- Набор данных должен быть опубликован в 2022 году и содержать **> 7 000 записей** о встречах видов
- Территория: Россия, Украина, Беларусь, **Казахстан**, Кыргызстан, Узбекистан, Таджикистан, Туркменистан, Молдова, Грузия, Армения, Азербайджан

<https://www.gbif.org/news/B00TTCi85Yly3qNo5KXSM/call-for-data-papers-describing-datasets-from-northern-eurasia-extended>

GBIF | Global Biodiversity Information Facility

Free and open access to  
biodiversity data

OCCURRENCES

SPECIES

DATASETS

PUBLISHERS

RESOURCES

Search



What is GBIF?

About GBIF Kazakhstan

*Periophthalmus barbarus* (Linnaeus, 1766) observed in Kubuneh, Gambia by supergan (CC BY-NC-SA 4.0)



2,246,447,878

Occurrence records



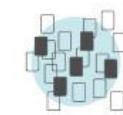
79,029

Datasets



1,926

Publishing institutions



7,894

Peer-reviewed papers  
using data

Глобальная Информационная Система о Биоразнообразии GBIF.org

GBIF.ORG

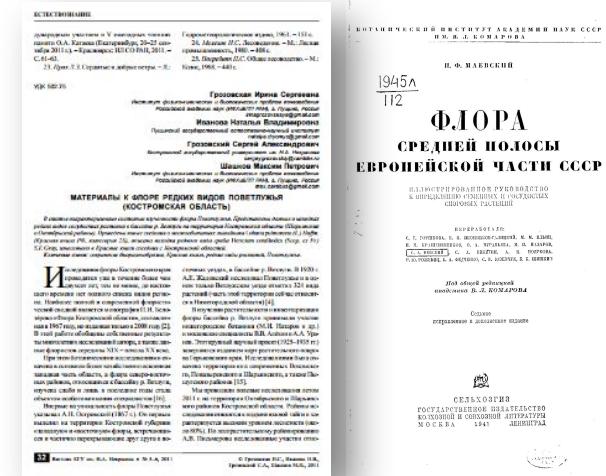
GBIF



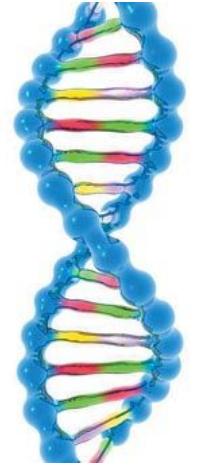
## Данные о коллекционных образцах



## Данные об окаменелостях



## Литературные данные



## Биоматериалы



## Наблюдения



## Данные автоматических регистраторов



## Данные площадных и маршрутных учетов

# Какие данные доступны через GBIF

# Типы наборов данных, доступных через GBIF

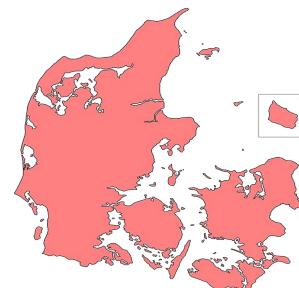
## Метаданные (METADATA only)



### Таксономический список CHECKLIST

для публикации  
таксономических данных:  
списки охраняемых  
видов, тематические  
видовые списки и др.

1 строка – 1 таксон



### Найдены OCCURRENCE DATA

Найдена - простое полевое  
наблюдение или  
коллекционный образец

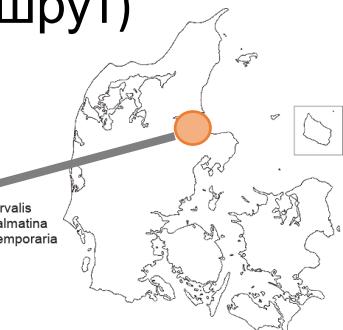
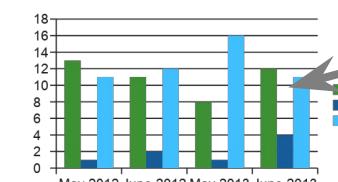
1 строка = 1 особь или 1 группа  
особей



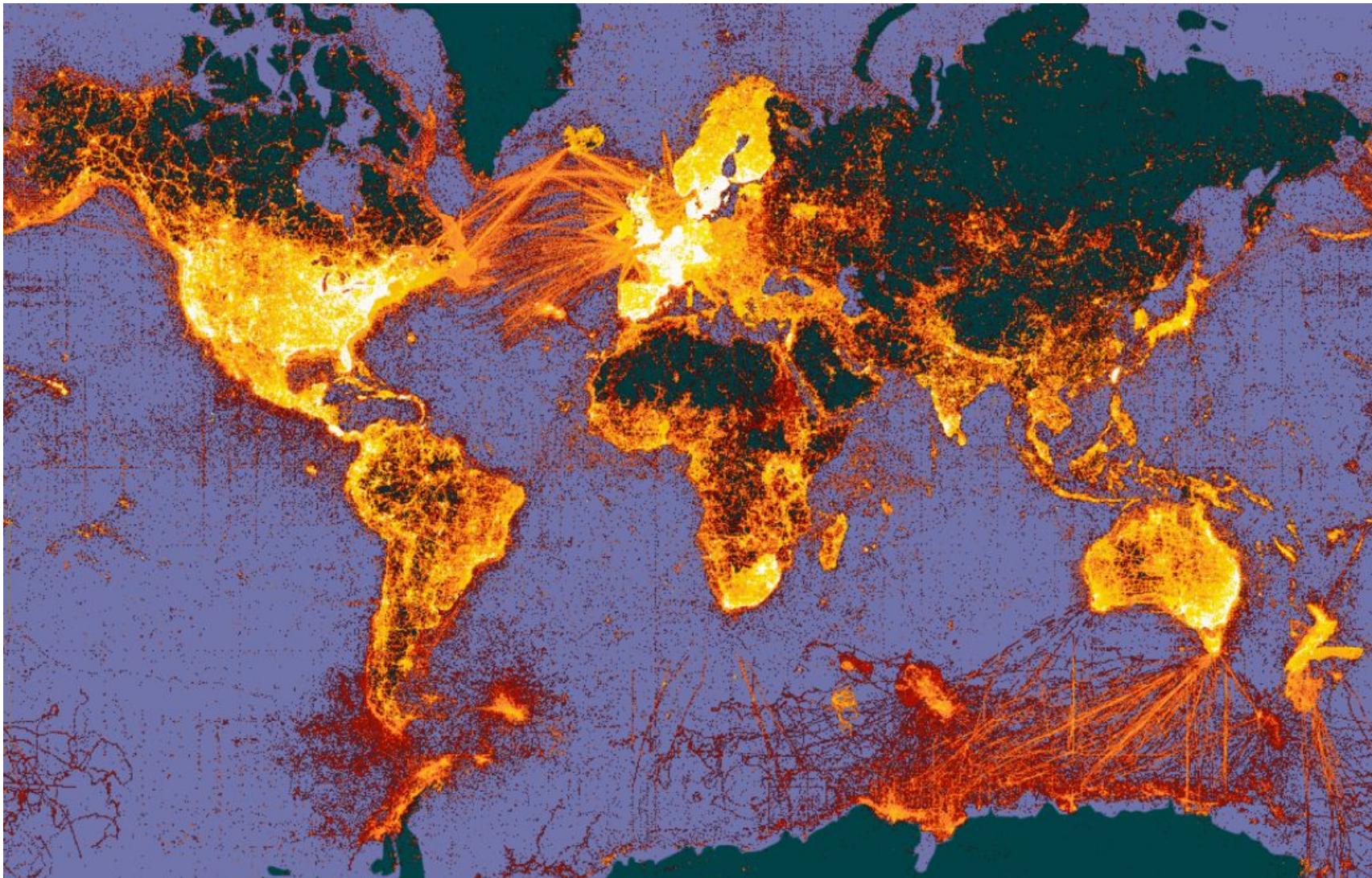
### Описания, учеты или сборы SAMPLE EVENT DATA

2 таблицы: данные о сборе  
+ данные о находках

1 запись на листе event =  
1 событие (площадка,  
 маршрут)



# Карта глобального биоразнообразия GBIF



# Публикация данных в GBIF

Google Sheets screenshot showing a data example for the BID CE Workshop 1 - SESSION 08. The sheet contains 15 rows of data with columns for Cat. Numb., University, Collector, No. of spec., YE, MO, DA, and Co. The data is from the University of Guatemala, with collectors including Betancur J, Fonnegra R, and Vargas I, and years ranging from 1991 to 1997.

	A	B	C	D	E	F	G
1	Cat. Numb.	University	Collector	No. of spec.	YE	MO	DA
2	UWP:100217	University of Guatemala	Betancur J	1	1991	5	11
3	UWP:100218	University of Guatemala	Betancur J	1	1991	5	11
4	UWP:101378	University of Guatemala	Fonnegra R	1	1994	5	31
5	UWP:101717	University of Guatemala	Betancur J	1	1993	11	9
6	UWP:101737	University of Guatemala	Betancur J	1	1993	11	7
7	UWP:102143	University of Guatemala	Betancur J	1	1994	4	20
8	UWP:102144	University of Guatemala	Betancur J	1	1994	4	20
9	UWP:102233	University of Guatemala	Vargas I	1	1995	12	6
10	UWP:103108	University of Guatemala	Cardona F	1	1996	4	3
11	UWP:104139	University of Guatemala	Fonnegra R	1	1996	5	15
12	UWP:104512	University of Guatemala	Callejas R	1	1995	7	21
13	UWP:105292	University of Guatemala	Acevedo P	1	1994	9	7
14	UWP:106768	University of Guatemala	Vargas I	1	1997	4	18
15	UWP:107441	University of Guatemala	Gonzales G	1	1997	4	12

GBIF dataset page for Belgian IFBL Flora Checklists (1939-1971). The page includes a map of Europe with a cluster of red dots representing the data distribution in Belgium, and a detailed description of the dataset.

**OCCURRENCE DATASET | REGISTERED 4 MAY 2010**

## Belgian IFBL Flora Checklists (1939-1971)

Published by [Botanic Garden Meise](#)  
by [Wouter Van Landuyt](#) • [Nicolas Noe](#)

**DATASET** **METRICS** **ACTIVITY** **DOWNLOAD** [DATASET HOMEPAGE](#)

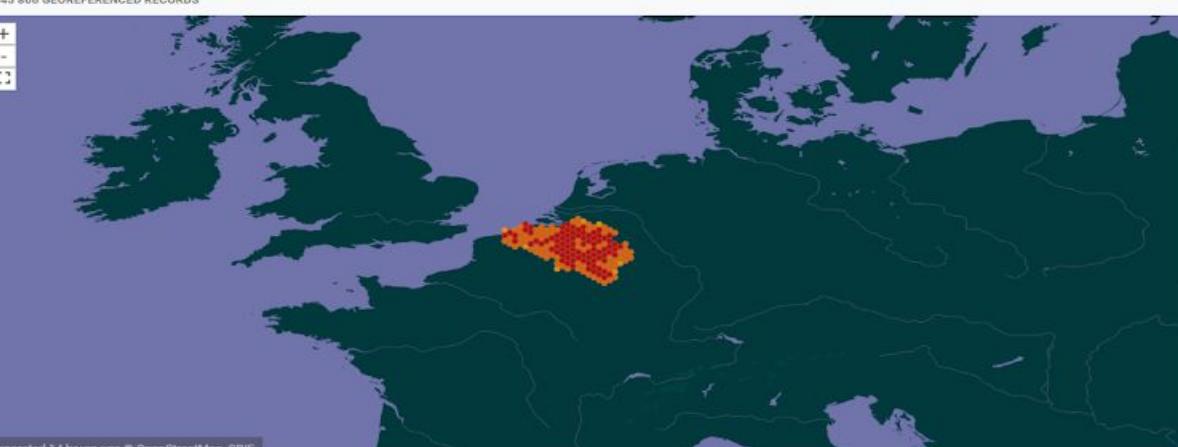
**1 045 806 OCCURRENCES** **9 CITATIONS**

The data in IFBL 1, 2 & 3 covers all of the IFBL 1 km<sup>2</sup> flora checklists sampled between 1939 and 1971. About 10000 original lists corresponding with some 1 200 000 data representative of the former distribution of vascular plant species in Belgium, were digitised. The IFBL data is Integrated in existing national and regional flora databases and will contribute to the realisation of regional Flora Atlases. The analysis of the digitised data will improve the possibilities to compare floral data over time. [more](#)

**Metadata Last Modified:** 5 March 2018  
**Data Last Changed:** 9 September 2015  
**License:** CC0 1.0  
[How to cite](#) [DOI](#) 10.15468/xmlbke

**1 045 806 Occurrences** **99.9% With taxon match** **100% With coordinates** **91% With year**

**1 045 806 GEOFERENCED RECORDS**



Generated 14 hours ago © OpenStreetMap, GBIF.

**Description** **Temporal** **Geographic**

The data in IFBL 1, 2 & 3 covers all of the IFBL 1 km<sup>2</sup> flora checklists sampled between 1939 and 1971. About 10000 original lists corresponding with some 1 200 000 data representative of the former distribution of vascular plant species in Belgium, were digitised. The IFBL data is integrated in existing national and regional flora databases and will contribute to the realisation of regional Flora Atlases. The analysis of the digitised data will improve the possibilities to compare floral data over time.



PUBLISHER | SINCE AUGUST 8, 2022

# Institute of Zoology of the Republic of Kazakhstan

ABOUT

METRICS

HOME PAGE

110 OCCURRENCES

1 DATASET

**Description:** Mission: To develop Kazakhstan zoology science from animal world investigations - from Republic of Kazakhstan to around of the globe. Vision: Institute of Zoology RK - leading world class science organization of fundamental and practical zoology. Institute of Zoology RK main directions: Kazakhstan wildlife investigation; animal population dynamics changes from environmental impact studying; studying of evolution, phylogeny, taxonomy diversity of present and past wildlife; ecologically valuable zoology problem solving.



**Endorsed by:** Participant Node Managers Committee

**Administrative contact:** Roman Jashenko

**Country or area:** Kazakhstan



## CONTACTS

Institute of Zoology of the Republic of Kazakhstan  
Al-Farabi, 93  
Almaty A15E3K2  
Kazakhstan  
[info@zool.kz](mailto:info@zool.kz)  
[+77059108472](tel:+77059108472)  
<https://zool.kz/eng/main-page/>

Vitale Morozov  
Point of contact  
[vitaliy.morozov@zool.kz](mailto:vitaliy.morozov@zool.kz)  
[+77059108472](tel:+77059108472)

Roman Jashenko  
Administrative point of contact  
[roman.jashenko@zool.kz](mailto:roman.jashenko@zool.kz)  
[+77017233525](tel:+77017233525)

SAMPLING EVENT | REGISTERED JULY 25, 2017

The communities of terrestrial macrofungi in different forest types of boreal zone in West Siberia

Published by [Yugra State University Biological Collection \(YSU BC\)](#)

Nina Filippova

DATASET METRICS ACTIVITY DOWNLOAD HOME PAGE

5,693 OCCURRENCES 4 CITATIONS

The dataset includes the results of a long-term monitoring program for larger fungi fruiting using permanent plots located in different forest types in central part of Western Siberia (nearby Khanty-Mansiysk). Our goal was to describe the quantitative and qualitative structure and spatial variation of the community of macromycetes, show its dynamics seasonally and inter-annually and also elucidate the relationship between the fruiting and climate variables. A total of 320 circular 5 m<sup>2</sup> (for a ... [More](#)



Metadata last modified: August 8, 2019

Data last changed: August 8, 2019

Hosted by: Institute of Mathematical Problems of Biology, Russian Academy of Sciences

License: CC BY 4.0

[How to cite](#) DOI [10.15468/ge1hkl](https://doi.org/10.15468/ge1hkl)

5,693 Occurrences

100% With taxon match

100% With coordinates

100% With year

5,693 GEOFERENCED RECORDS



## Citation

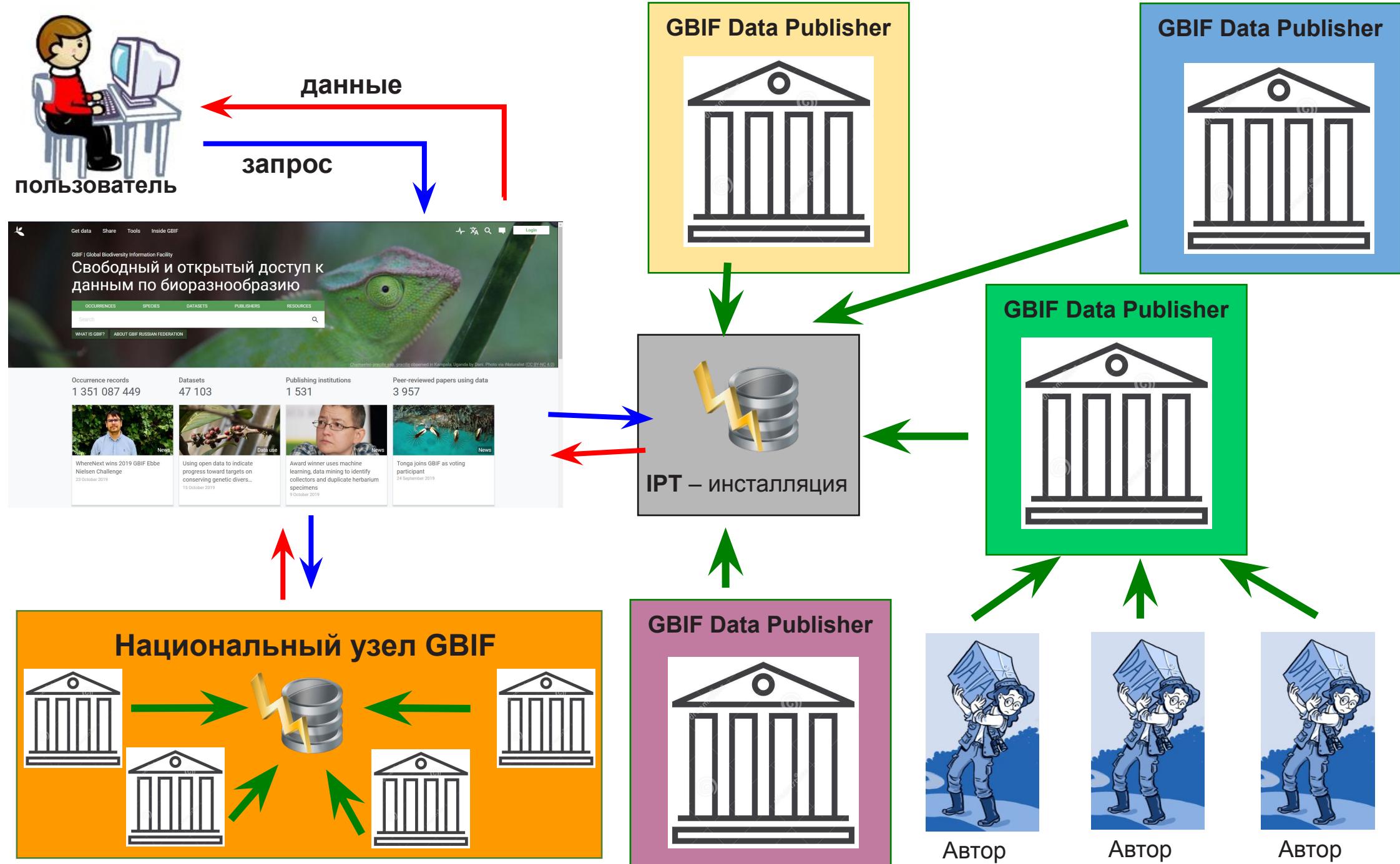
Filippova N (2018). The communities of terrestrial macrofungi in different forest types of boreal zone in West Siberia.

Version 1.2. Yugra State University Biological Collection (YSU BC). Sampling event dataset

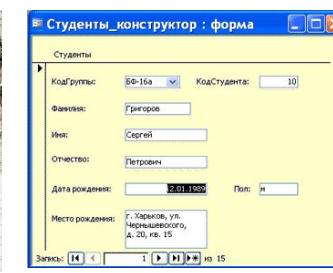
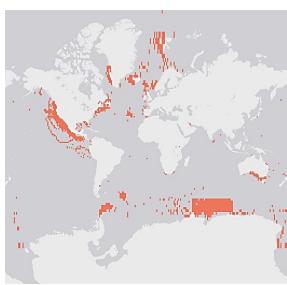
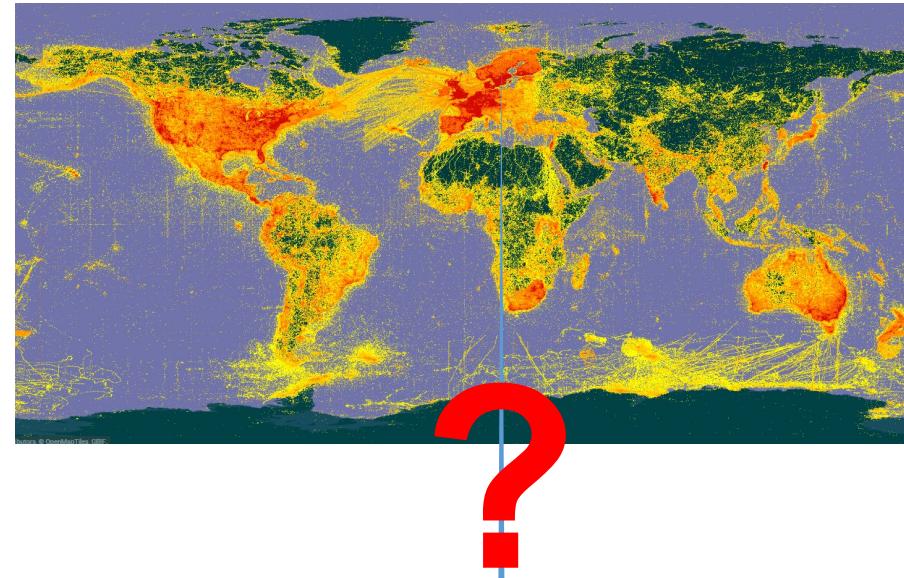
<https://doi.org/10.15468/ge1hkl> accessed via GBIF.org on 2019-03-22.

# Набор данных (dataset), доступный через GBIF

# Как устроена сеть GBIF



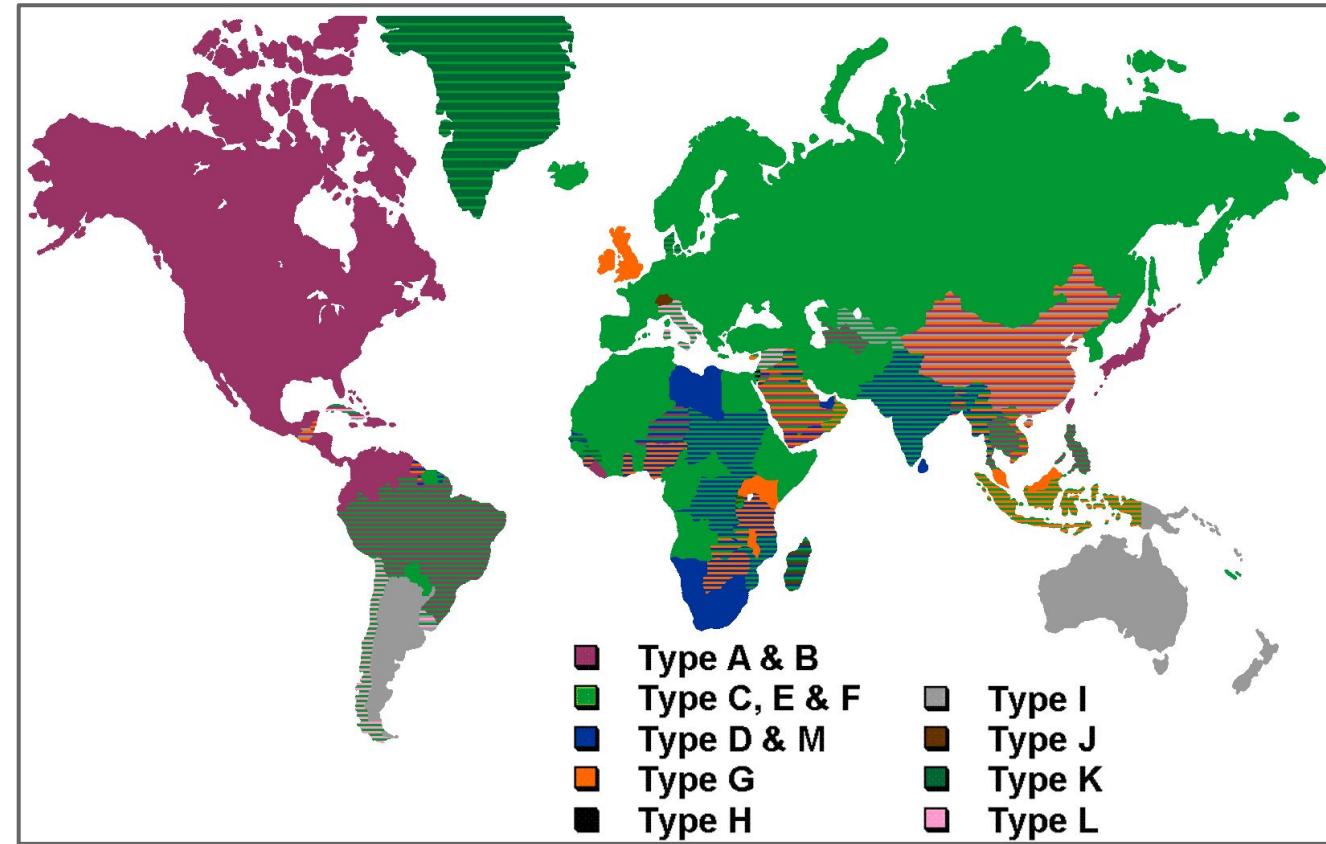
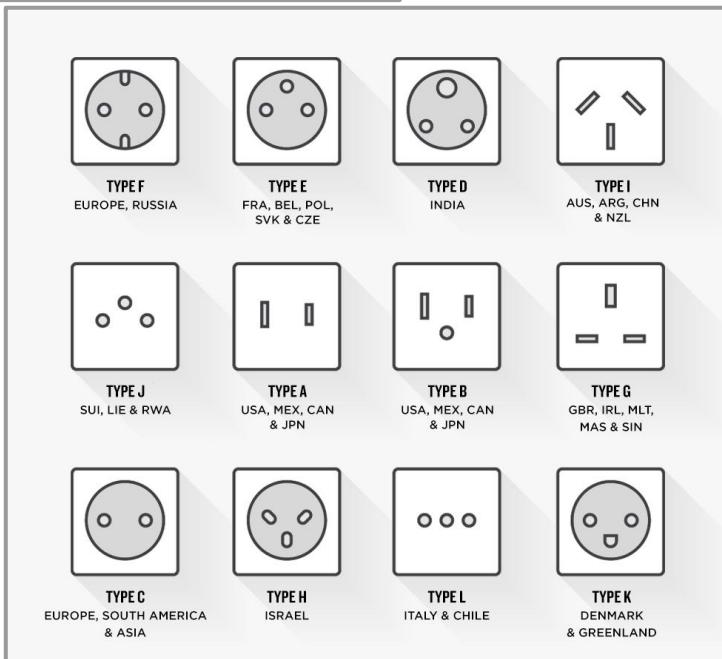
# Как объединить данные в глобальную систему?

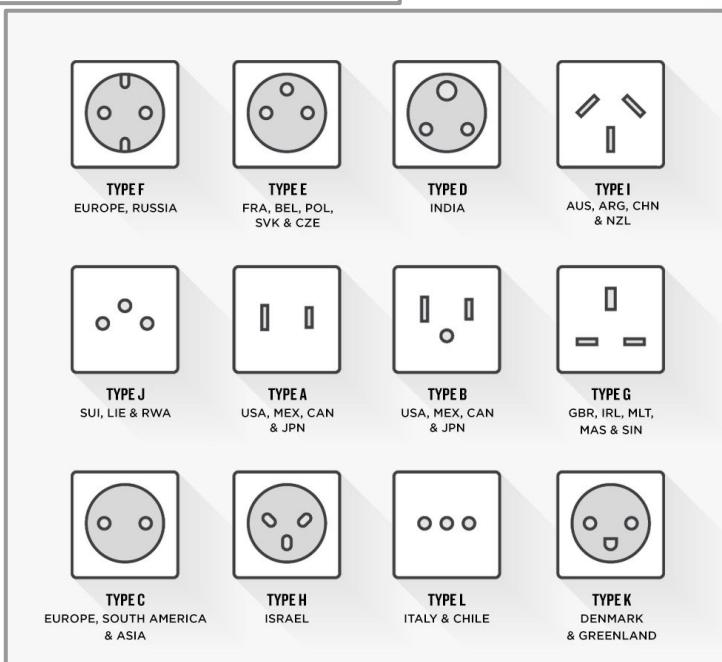


№	ФИО	Курсант	Примеч.	Документ.
1	Григорьев А.М.	Радиотехн.		400
2	Иванов И.М.	Радиотехн.		400
3	Петров П.И.	Механическ.		400
4	Маркович Ф.Л.	Радиотехн.		400
5	Сергеев С.С.	Механическ.		400
6	Петрович В.В.	Вертолет.		400
7	Григорьев В.В.	Вертолет.		400
8	Григорьев А.М.	Огнеупр.		400
9	Петрович В.В.	Огнеупр.		400
10	Сергеев С.С.	Вертолет.		400
11	Петров С.С.	Вертолет.		400
12	Иванов С.С.	Радиотехн.		400
13	Петрович С.С.	Механическ.		400
14	Маркович С.С.	Радиотехн.		400
15	Григорьев С.С.	Вертолет.		400
16	Петрович С.С.	Вертолет.		400
17	Сергеев С.С.	Огнеупр.		400
18	Петров С.С.	Огнеупр.		400
19	Иванов С.С.	Огнеупр.		400
20	Петрович С.С.	Механическ.		400
21	Григорьев С.С.	Радиотехн.		400
22	Иванов С.С.	Радиотехн.		400
23	Петров С.С.	Механическ.		400
24	Маркович С.С.	Радиотехн.		400
25	Григорьев С.С.	Вертолет.		400



ИСТОЧНИКИ ДАННЫХ





# Darwin Core – основной стандарт данных в GBIF

## Darwin Core

Darwin Core is a standard maintained by the [Darwin Core maintenance group](#). It includes a glossary of terms intended to **facilitate the sharing of information about biological diversity** by providing identifiers, labels, and definitions. Darwin Core is primarily based on taxa, their occurrence in nature as documented by observations, specimens, samples, and related information.

[technical specification](#)[current standard](#)[2009](#)[Darwin Core website !\[\]\(1c70f21f694d12b9fc928edb998ea27b\_img.jpg\)](#)[Find us on GitHub !\[\]\(179f6b45bf59a6537d86a5664856855b\_img.jpg\)](#)<http://rs.tdwg.org/dwc/terms/>

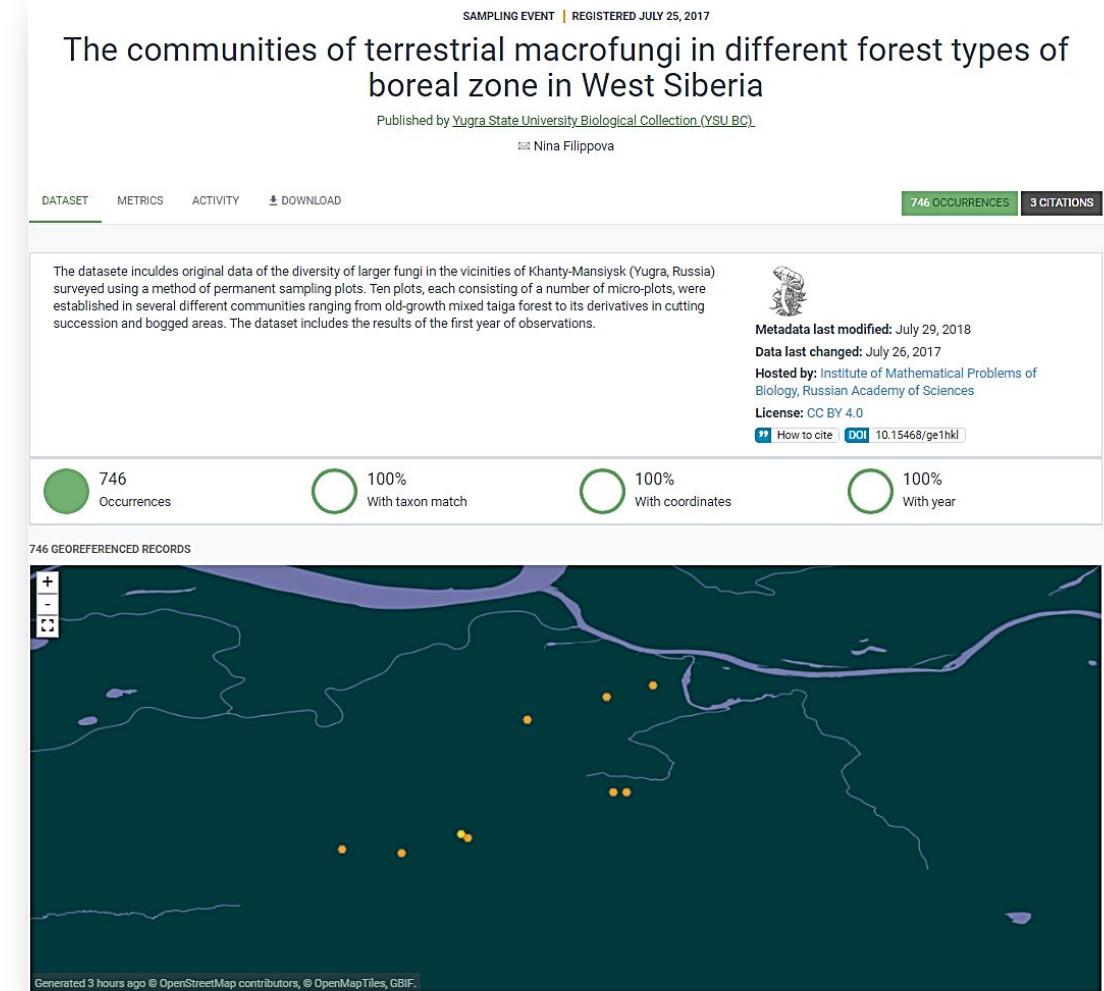
Система обмена информацией в GBIF построена на открытых стандартах, центральный из которых – **Darwin Core (DwC)**, стандарт, разработанный специально для хранения данных по биоразнообразию.

**Стандарт** – набор терминов (заголовки столбцов в таблице с данными) и правила использования этих терминов.

# ЗАЧЕМ публиковать свои данные через глобальные порталы?



или



# ЗАЧЕМ публиковать свои данные через глобальные порталы?



- Проверяемость результатов научных исследований
- Закрепление авторства собственных данных через DOI
- Повышение обнаружимости данных и исследований
- Возможности совместных исследований

- Повышение научного использования коллекций
- Лучшая сохранность фондовых образцов
- Систематизация и улучшение качества данных



- Выполнение обязательств в рамках международных Конвенций.
- Достижение Целей в области устойчивого развития, в том числе связанных с здоровьем человека, продовольственной безопасностью и глобальным изменением климата, а также сохранением биоразнообразия на суше и в океане.

Файл Главная Вставка Разметка страницы Формулы Данные Рецензирование Вид Надстройки Что вы хотите сделать? Вход Общий доступ

Сортировка и фильтр Выделить

Сортировка Найти и фильтр

Редактирование

B2 Achipteria coleoptrata (Linnaeus, 1758)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
5	iee-2010-04	Achipteria coleoptrata (Linnaeus, 1758)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.0568	30.6463	WGS1984	10	5	Ind / soil caspен for A.S.	Zaltsev A.	IEE RAS				
6	iee-2010-05	Galluma obvia (Berlese, 1914)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.0993	30.3815	WGS1984	10	3	Ind / soil c bog	A.S.	Zaltsev A.	IEE RAS			
7	iee-2010-06	Tectocephus velutus (Michael 1880)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.0993	30.3815	WGS1984	10	1	Ind / soil c Bottomal A.S.	Zaltsev A.	IEE RAS				
8	iee-2010-07	Achipteria coleoptrata (Linnaeus, 1758)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.1019	30.3891	WGS1984	10	25	Ind / soil c mixed for A.S.	Zaltsev A.	IEE RAS				
9	iee-2010-08	Mediopelta hypographa (Mahunka 1987)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.1032	30.3907	WGS1984	10	52	Ind / soil c peatbog	A.S.	Zaltsev A.	IEE RAS			
10	iee-2010-09	Scheloribates laevigatus (Koch, 1835)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.4417	30.6801	WGS1984	10	36	Ind / soil c meadow	A.S.	Zaltsev A.	IEE RAS			
11	iee-2010-10	Microtritia minima (Berlese, 1904)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.4417	30.6801	WGS1984	10	26	Ind / soil c meadow	A.S.	Zaltsev A.	IEE RAS			
12	iee-2010-11	Rhysotritia duplicita (Grandjean 1953)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.4417	30.6801	WGS1984	10	4	Ind / soil c meadow	A.S.	Zaltsev A.	IEE RAS			
13	iee-2010-12	Scheloribates laevigatus (Koch, 1835)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.102	30.4306	WGS1984	10	21	Ind / soil c mixed for A.S.	Zaltsev A.	IEE RAS				
14	iee-2010-13	Parakalumnidae	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.057	30.6418	WGS1984	10	15	Ind / soil c mixed for A.S.	Zaltsev A.	IEE RAS				
15	iee-2010-14	Platynothrus peltifer (Koch, 1840)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.057	30.6418	WGS1984	10	20	Ind / soil c mixed for A.S.	Zaltsev A.	IEE RAS				
16	iee-2010-15	Xenilus tegeocranus (Hermann 1804)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.057	30.6418	WGS1984	10	2	Ind / soil c mixed for A.S.	Zaltsev A.	IEE RAS				
17	iee-2010-16	Hoplophirhacarus Illinoisensis (Ewing, 1909)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.171	30.6404	WGS1984	10	24	Ind / soil c Transition A.S.	Zaltsev A.	IEE RAS				
18	iee-2010-17	Tectocephus velutus (Michael 1880)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.171	30.6404	WGS1984	10	3	Ind / soil c Transition A.S.	Zaltsev A.	IEE RAS				
19	iee-2010-18	Trichoribates trimaculatus (C.L.Koch, 1835)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.171	30.6404	WGS1984	10	1	Ind / soil c Transition A.S.	Zaltsev A.	IEE RAS				
20	iee-2010-19	Opiella nova (Oudemans, 1902)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.1042	30.39	WGS1984	10	28	Ind / soil c Raised pe A.S.	Zaltsev A.	IEE RAS				
21	iee-2010-20	Nanhermannia dorsalis (Banks 1902)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.1042	30.39	WGS1984	10	3	Ind / soil c Raised pe A.S.	Zaltsev A.	IEE RAS				
22	iee-2010-21	Phthiracarus globosus (Koch, 1835)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.1019	30.3891	WGS1984	10	1	Ind / soil c Raised pe A.S.	Zaltsev A.	IEE RAS				
23	iee-2010-22	Opiella nova (Oudemans, 1902)	SPECIES	2010	2010-06-0 RU	Bezhanits Polistovsk	57.0984	30.3809	WGS1984	10	2	Ind / soil c Raised pe A.S.	Zaltsev A.	IEE RAS				
24	iee-2010-23	Tectocephus velutus (Michael 1880)	SPECIES	2010	RU	Bezh	57.0984	30.3809	WGS1984	10	11	Ind / soil						
25	iee-2010-24	Zetomimus furcatus (Warburton & Pearce, 1959)	SPECIES	2010	2010-06-0 RU	Bezh	57.0984	30.3809	WGS1984	10	1	Ind / soil						
26	iee-2010-25	Hoplophirhacarus Illinoisensis (Ewing, 1909)	SPECIES	2010	2010-06-0 RU	Bezh	50.6452	57.1743	WGS1984	10	12	Ind / soil						
27	iee-2010-26	Scheloribates laevigatus (Koch, 1835)	SPECIES	2010	2010-06-0 RU	Bezh	57.1743	30.6452	WGS1984	10	1	Ind / soil						
28	iee-2010-27	Chamobates cuspidatus (F.	SPECIES								1	Ind / soil						
29	iee-2010-28	Scheloribates laevigatus (C.L.	SPECIES								6	Ind / soil						
30	iee-2010-29	Minuthozetes semirufus	SPECIES								4	Ind / soil						

Region vs Patient

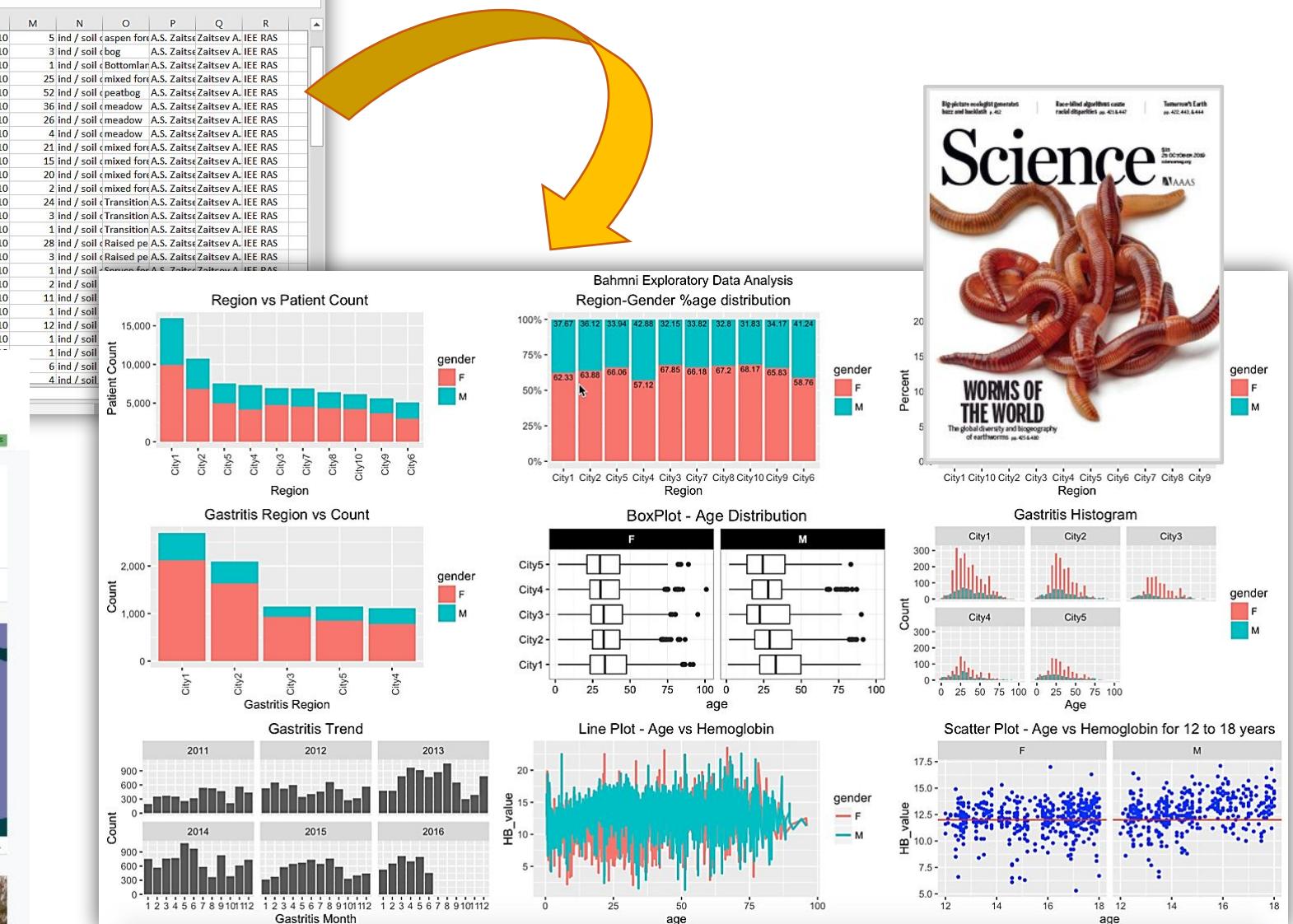
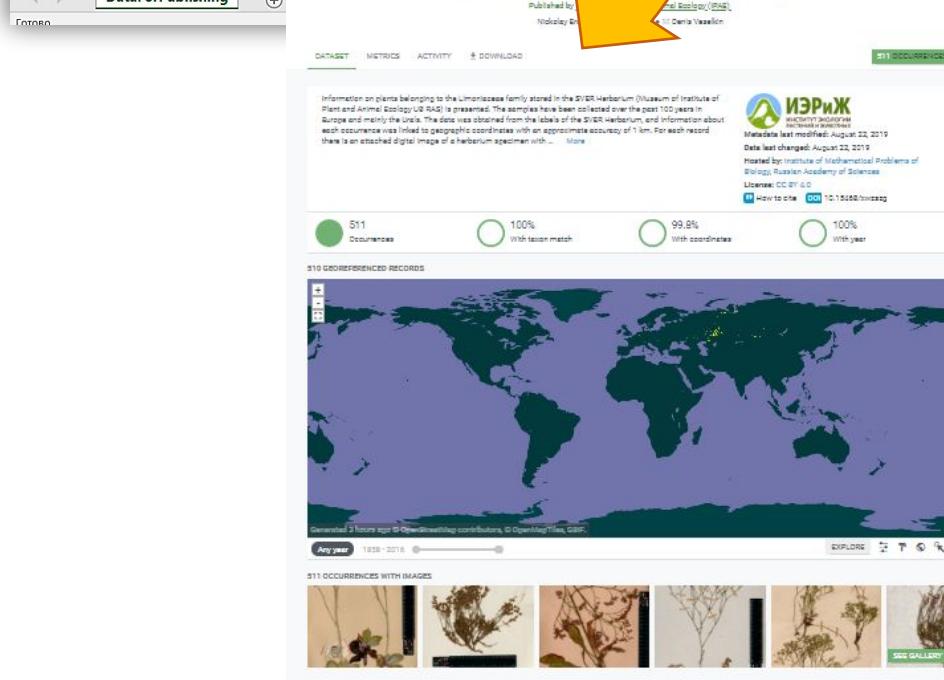
15,000

10,000

5,000

0

Patient Count



<https://i.pinimg.com/originals/5c/e1/ab/5ce1ab055f79a92c5c6a42586ca67de7.png>

# Зачем нужна проверка данных (data cleaning) ?

A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P
Registration number	Full name	Status	Conservation	Stage	males	Collector	Collection number	Date of collection	Determiner	Locality	Family	Class	Country	Latitude	Longitude
1,00	Eucidaris tribuloides africana Mortensen Ht	In alcohol		1		Dartevelle E.	132B	XI.1965.III-1966	Mortenser M'Vassa, Fausse Pointe No Cidaridae	Echinidae	Congo, D. R.				
2,00	Arbacia lixula africana Troschel	In alcohol		1		Dartevelle E.		VII.1938	Mortenser M'Vassa, Fausse Pointe No Arbacidae	Echinidae	Congo, D. R.				
3,00	Stomopneustes variolaris Lamarck			1		Burgeon L.		1934	Mortenser Dar es Salaam	Stomopneustidae	Echinidae	Tanzania	S 06° 48' E 00° 00'		
4,00	Arbacia lixula africana Troschel			4		Dartevelle E.		XI.1937	Mortenser Cabinda, Ponta n'Gelo	Arbacidae	Echinidae	Angola			
9,00				1		Dartevelle E.	B.4	XI.1937	Mortenser Cabinda	Cidaridae	Echinidae	Angola			
10,00	Arbaciella elegans Mortensen			2		Dartevelle E.		XI.1937	Mortenser Cabinda	Arbacidae	Echinidae	Angola			
18,00	Arbacia lixula africana Troschel			1		Dartevelle E.	G.I	VII.1938	Mortenser M'Vassa, Fausse Pointe No Arba	Echinidae	Echinidae				
19,00	Paracentrotus lividus Lamarck			4		Dartevelle E.		I.1937	Mortenser Lobito bay	Echinidae	Echinidae				
23,00	Arbacia lixula africana Troschel			1		Dartevelle E.		XI.1937	Mortenser Cabinda	Arbacidae	Echinidae				
26,00	Echinometra lucunter L.	Microscop		1		Dartevelle E.	B.4	VII.1938	Mortenser M'Vassa, Fausse Pointe No Echinid	Echinidae	Echinidae				
27,00	Eucidaris tribuloides africana Mortensen			1		Dartevelle E.		VII.1938	Mortenser Loango, Pointe	Echinidae	Echinidae				
28,00	Astropecten gruveli Kochler			1		Dartevelle E.		I.1937	Mortenser Lobito bay	Echinidae	Echinidae				
29,00	Astropecten michaelensi Kochler			1		Dartevelle E.		XI.1937	Mortenser Cabinda	Echinidae	Echinidae				
31,00	Asterina sp.			1		Dartevelle E.		Lobito	Mortenser Lobito Bay	Echinidae	Echinidae				
32,00	Astropecten dahomense L.			1		Dartevelle E.			Banana	Echinidae	Echinidae				
35,00	Radiorotula orbicularis radiata			1		Dartevelle E.		I.1937	Dartevelle Lobito Bay	Echinidae	Echinidae				
53,00	Radiorotula orbicularis L.			17		Vrydagh		1937	Dartevelle Lobito Plage	Echinidae	Echinidae				
78,00	Radiorotula orbicularis L.			6		Dartevelle E.		1937	Dartevelle Lobito Bay	Echinidae	Echinidae				
170,00	Radiorotula orbicularis L.			35J		Dartevelle E.		1933	Dartevelle Lobito Bay	Echinidae	Echinidae				
206,00	Radiorotula orbicularis L.			40J		Dartevelle E.		I.1937	Dartevelle Lobito Bay	Echinidae	Echinidae				
246,00	Radiorotula orbicularis L.			1		Dartevelle E.	106	20-27.II.1959	Dartevelle Lobito Bay	Echinidae	Echinidae				
247,00	Rotula angusti Klein			1		Dartevelle E.			Dartevelle Banane	Scutellidae	Echinidae	Congo, D. R.	S 06° 00' E 012° 24'		
248,00	Rotula angusti Klein			1		Bitremieuy R.G.		1930	Dartevelle Moanda, Bas-Congo	Scutellidae	Echinidae	Congo, D. R.	S 05° 56' E 012° 21'		
347,00	Radiorotula orbicularis L.			14		Dartevelle E.		1933	Dartevelle Lobito Bay	Scutellidae	Echinidae	Angola			
362,00	Radiorotula orbicularis L.			6J		Dartevelle E.		1933	Dartevelle Lobito Bay	Scutellidae	Echinidae	Angola			
419,00	Radiorotula orbicularis L.			1		Dartevelle E.		1933	Dartevelle Lobito Bay	Scutellidae	Echinidae	Angola			
420,00	Paracentrotus lividus Lamarck	Pt	In alcohol	1		Dartevelle E.		VI.1938	Mortenser M'Vassa, fausse Pointe no Echinidae	Echinidae	Echinidae				
424,00	Arbacia lixula africana Troschel		In alcohol	1		Dartevelle E.		X.1937	Mortenser Landana	Arbacidae	Echinidae				
425,00	Arbaciella elegans Mortensen		In alcohol	1		Dartevelle E.	13 P 10	X.1937	Mortenser Landana	Arbacidae	Echinidae				
426,00	Paracentrotus lividus Lamarck		In alcohol	1		Dartevelle E.		VI.1938	Mortenser M'Vassa, Fausse Pointe no Echinidae	Echinidae	Echinidae				
427,00	Astropecten michaelensi Kochler		In alcohol	1		Etienne Dr.		3.V.1910	Mortenser Banane	Astropectinidae	Astropectinidae	Congo, D. R.	S 06° 00' E 012° 24'		
428,00	Astropecten michaelensi Kochler		In alcohol	1		Schouteden Dr.		X.1930	Mortenser Banane	Astropectinidae	Astropectinidae	Congo, D. R.	S 06° 00' E 012° 24'		
430,00	Ophiactis africana Koehler		In alcohol	1		Dartevelle E.		VI-IX.1956	Mortenser Vista, Bas Congo	Ophiactidae	Ophiactidae	Congo, D. R.	S 05° 52' E 012° 17'		
431,00	Ophiactis africana Koehler		In alcohol	1		Dartevelle E.		II.1938	Mortenser Vista, Bas Congo	Ophiactidae	Ophiactidae	Congo, D. R.	S 05° 52' E 012° 17'		
432,00			In alcohol	1		Dartevelle E.		X.1937	Landana	Ophiactidae	Ophiactidae				
433,00	Radiorotula orbicularis L.	Other		1		Dartevelle E.			Banane	Scutellidae	Echinidae	Congo, D. R.	S 06° 00' E 012° 24'		
434,00	Radiorotula orbicularis L.			2		Dartevelle E.		I.1937	Lobito Bay	Scutellidae	Echinidae	Angola			
438,00	Radiorotula			1		Dartevelle E.		1933	Lobito Bay	Scutellidae	Echinidae	Angola			

Даты в разных форматах

Лишние ряды

Пустые ячейки

Данные всегда  
содержат  
ошибки!

Числа в разных  
форматах

# Качество данных

относительная концепция, которая зависит от способа использования этих данных



# Наиболее распространенные ошибки в данных

## Технические ошибки:

опечатки, пропущенные значения, лишние пробелы, корректность диапазонов для дат, соответствие типа данных полю, в котором они содержатся

## Ошибки формата данных

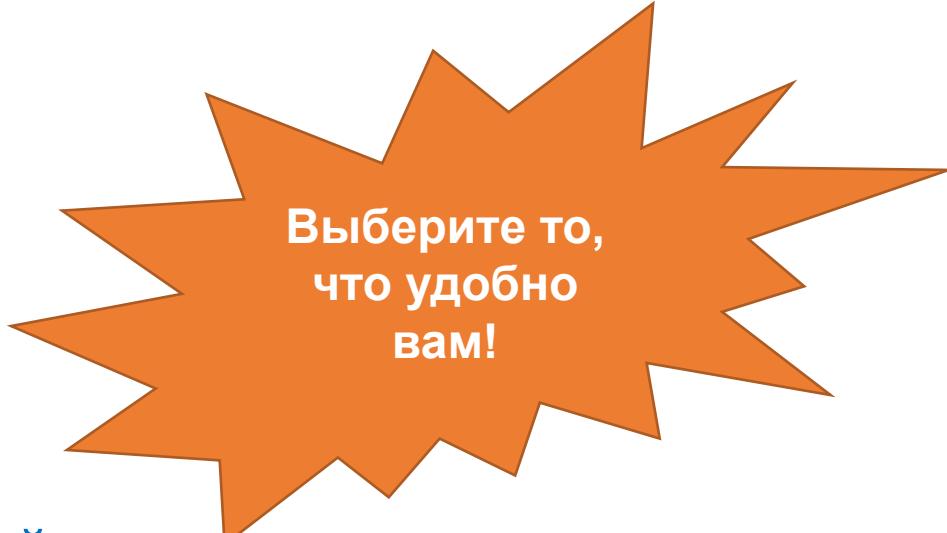
**Согласованность данных:** соответствие даты сбора, идентификации, обновления и оцифровки, координаты всех точек находятся в указанном регионе, точки находок сухопутных видов находятся на суше и т.д.

**Номенклатурные ошибки:** соответствие названия таксонов выбранному справочнику, соответствие других значений справочным

# Инструменты для поиска и исправления технических ошибок и ошибок в данных

## Текстовые редакторы:

- BBEdit (Mac)
- Notepad++ (Windows)
- Emacs, Vi (Unix, Linux)



Выберите то,  
что удобно  
вам!

## Инструменты, работающие с командной строкой:

- R
- RStudio (графический пользовательский интерфейс)
- Скрипты: Perl, Python,...



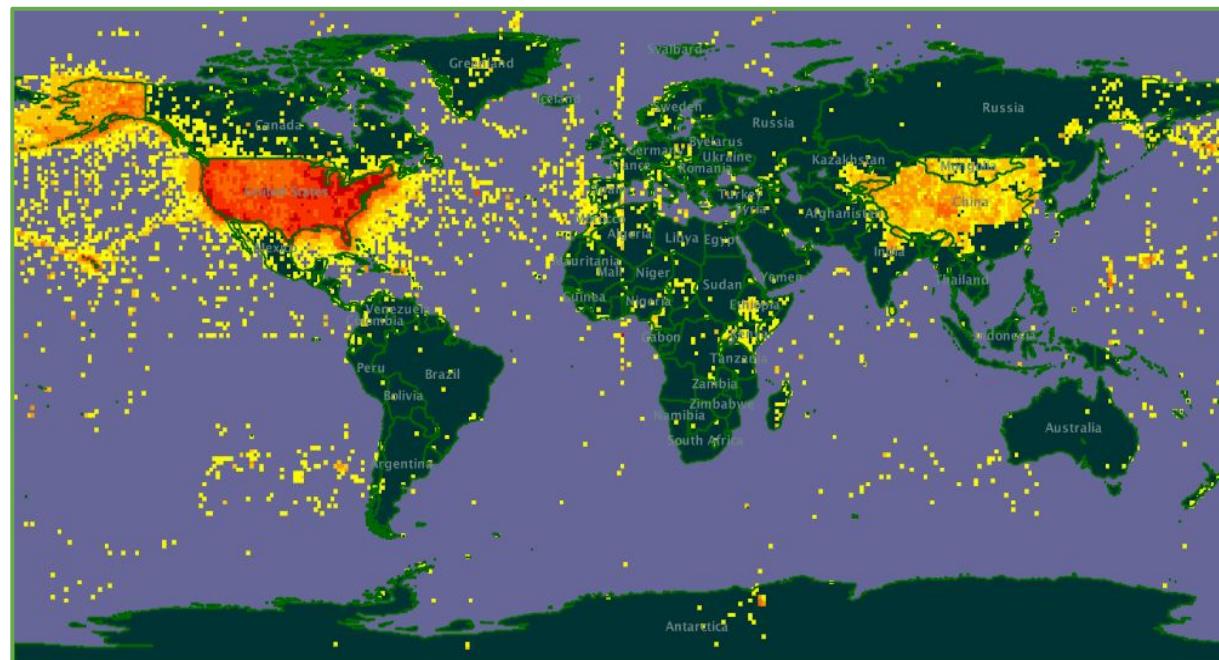
Excel

Refine OPEN 

QGIS

# Ошибки представления пространственных данных

- Широта и долгота перепутаны местами
  - Неправильно указано полушарие
  - Нулевые значения
  - Неизвестная система координат
  - Ошибки преобразования координат из одной системы в другую или из одной формы представления в другую



# Инструменты для верификации пространственных данных: QGIS

- Настольная (локальная) геоинформационная система (ГИС)
- Для трансформации, анализа, визуализации, проверки, верификации и т.д.
- <http://www.qgis.org>
- Open source!



RStudio

File Edit Code View Plots Session Build Debug Profile Tools Help

plots2022.R transsects2022deadwood.R Untitled1\*

Source on Save Run Source

```
1 setwd("~/Kologriv")
2 deadwood=read.csv('transects2022deadwood.csv')
3 dim(deadwood)
4 str(deadwood)
5 names(deadwood)
6
7 library('dplyr') #для подсчета частот
8 library('ggplot2')
9 library('tidyverse') #для создания перекрестной таблицы
10
11 ### число деревьев разных видов по трансектам
12 treescount=matrix(ncol=4)
13 colnames(treescount)=c('tree', 'count', 'numbertransect', 'site')
14 for (i in 1:15) {
15   a=count(deadwood[deadwood$transect==i], deadwood[deadwood$transect==i]$tree)
16   b=cbind(a, number=i, if (i<10) {site='core'} else {site='sekh'} )
17   names(b)=c('tree', 'count', 'numbertransect', 'site')
18   treescount=rbind(treescount, b)
19 }
20 treescount
21 treescount=treescount[2:45,]
22
```

49:61 (Top Level)

Console Terminal Jobs

```
~/Kologriv/ 
> dim(sekhacount)
[1] 10 4
> corecount$numbertransect=factor(corecount$numbertransect) # номер трансекты в фактор
> names(corecount)
[1] "tree"          "count"         "numbertransect" "site"
> ggplot(corecount, aes(fill=tree, y=count, x=numbertransect)) +
+   geom_bar(position='stack', stat="identity") +
+   scale_fill_manual(values = c('#1890F8', '#E07000', '#801088', '#D090A0', '#F8F800', '#402088', '#E810
8')) +
+   labs(x = 'Трансекты', y = 'Число стволов, шт.')+
+   guides(fill=guide_legend(title="деревья"))+
+   theme(axis.text=element_text(size=12),
+         axis.title=element_text(size=14,face="bold"))+
+   theme(legend.text = element_text(size=14), legend.title=element_text(size=16, face='bold'))
```

Environment History Connections

Global Environment

corecount	34 obs. of 4 variables
deadwood	353 obs. of 8 variables
monthly	695 obs. of 5 variables
monthly1	695 obs. of 3 variables
monthly13	565 obs. of 4 variables
monthly2	695 obs. of 3 variables
monthly3	695 obs. of 3 variables
monthly7	58 obs. of 5 variables
nearest	Formal class SpatialPointsDataFrame
occurs	3448 obs. of 6 variables
occursDistinct	1575 obs. of 6 variables

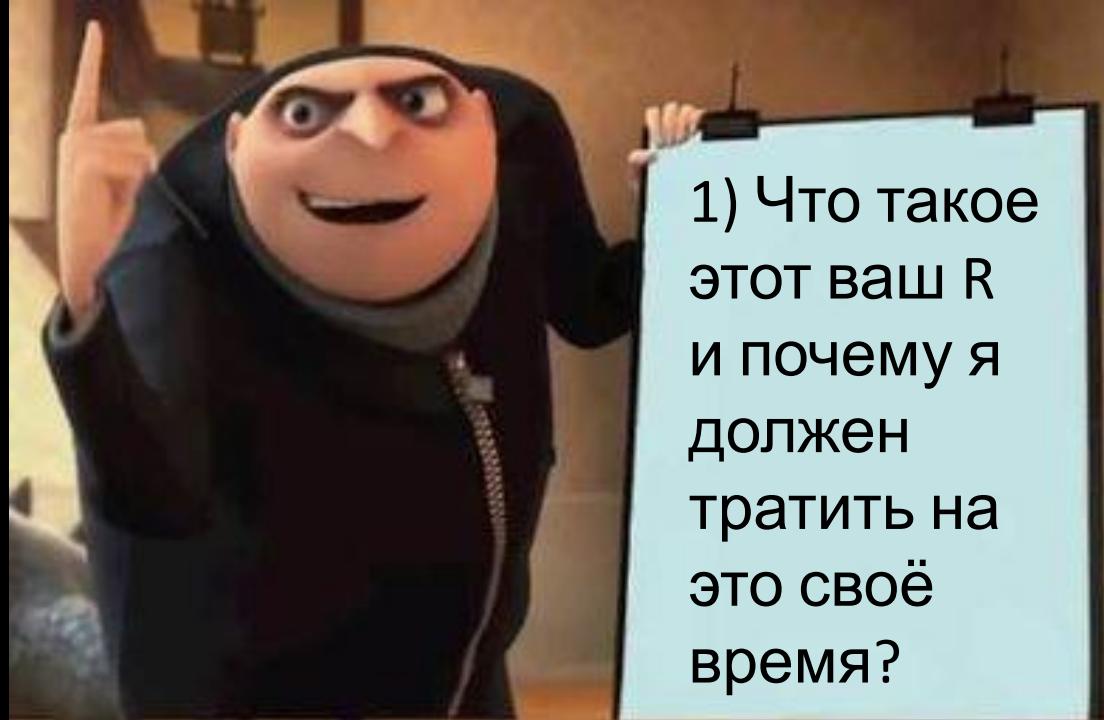
Files Plots Packages Help Viewer

Zoom Export

Деревья

- береза
- вяз
- ель
- клен
- липа
- пихта
- рябина

Трансекты	береза	вяз	ель	клен	липа	пихта	рябина	Всего
1	2	1	12	5	0	0	0	24
2	0	0	15	5	0	0	0	20
3	5	0	15	10	0	0	0	40
4	5	0	10	5	2	0	0	27
5	3	0	5	5	1	0	0	15
6	5	0	5	5	0	0	0	15
7	5	0	5	5	0	0	0	15
8	0	0	15	10	2	0	0	27
9	0	0	15	10	2	0	0	30



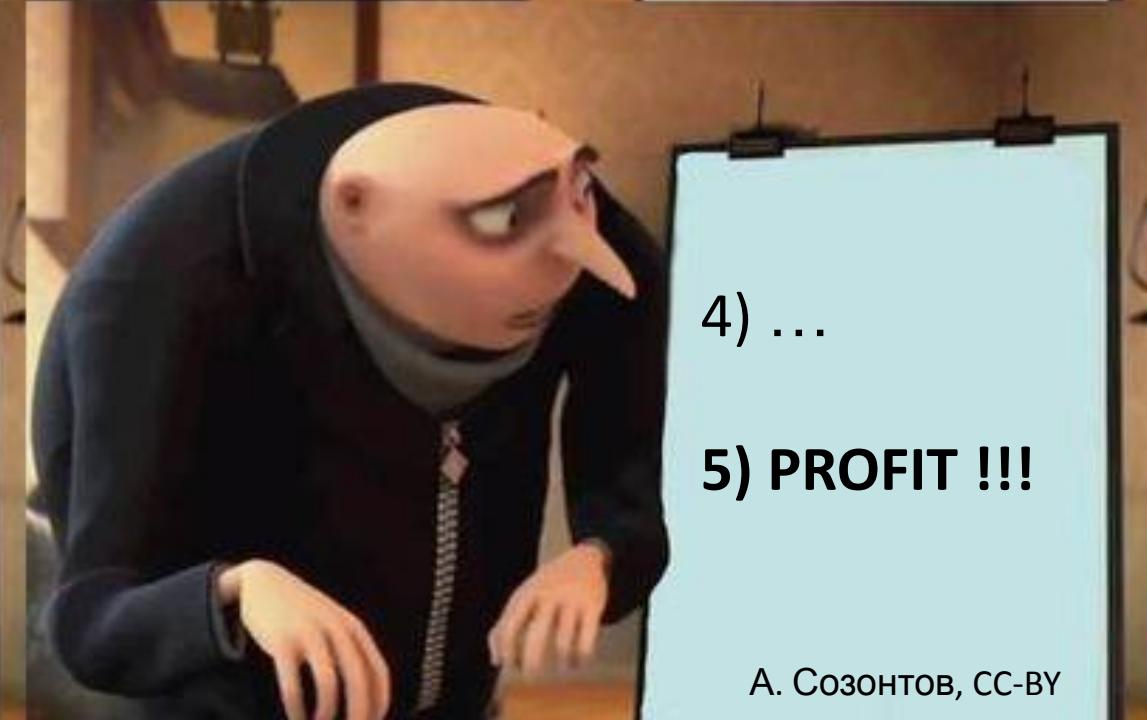
1) Что такое  
этот ваш R  
и почему я  
должен  
тратить на  
это своё  
время?



2) Чего R  
умеет такого,  
что я не могу  
сделать в  
Excel / Past /  
Statistica и  
другом ПО?



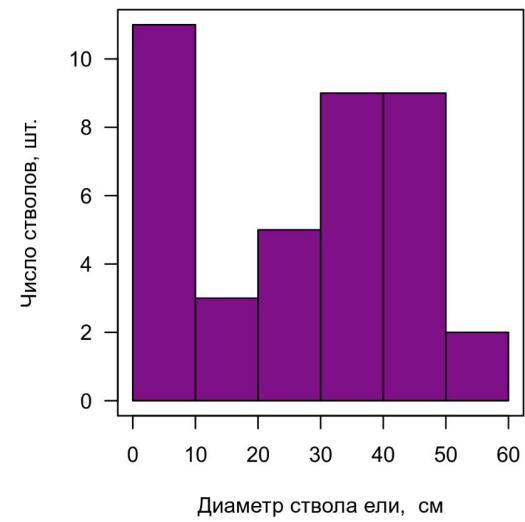
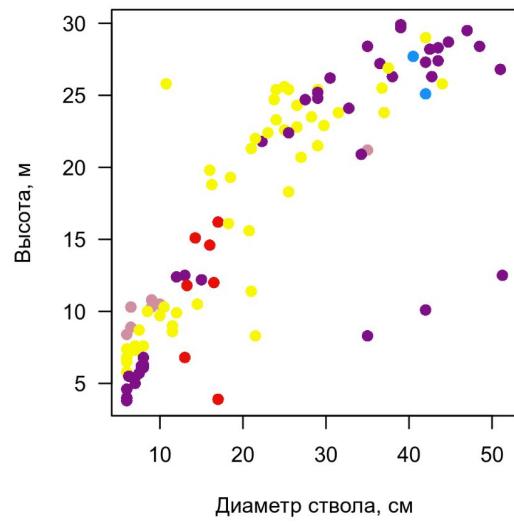
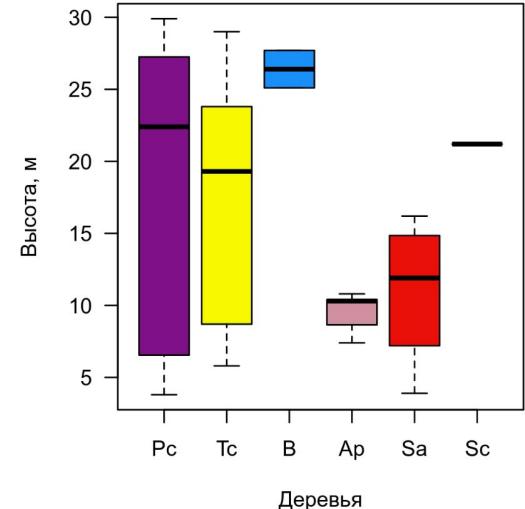
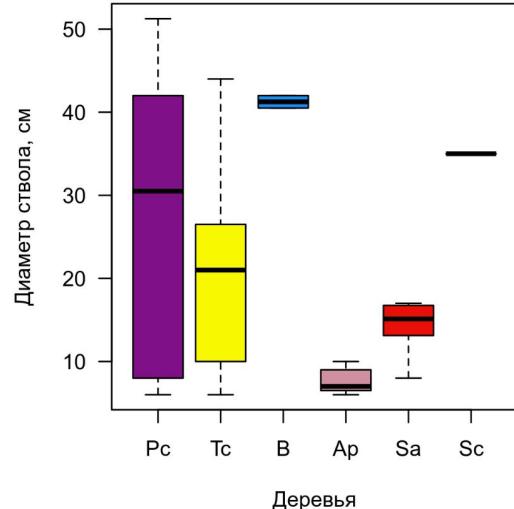
3) Считать это  
ладно, но  
сумеет ли R в  
два клика  
вывести мне  
на экран  
график, карту,  
диплом или  
диссертацию?

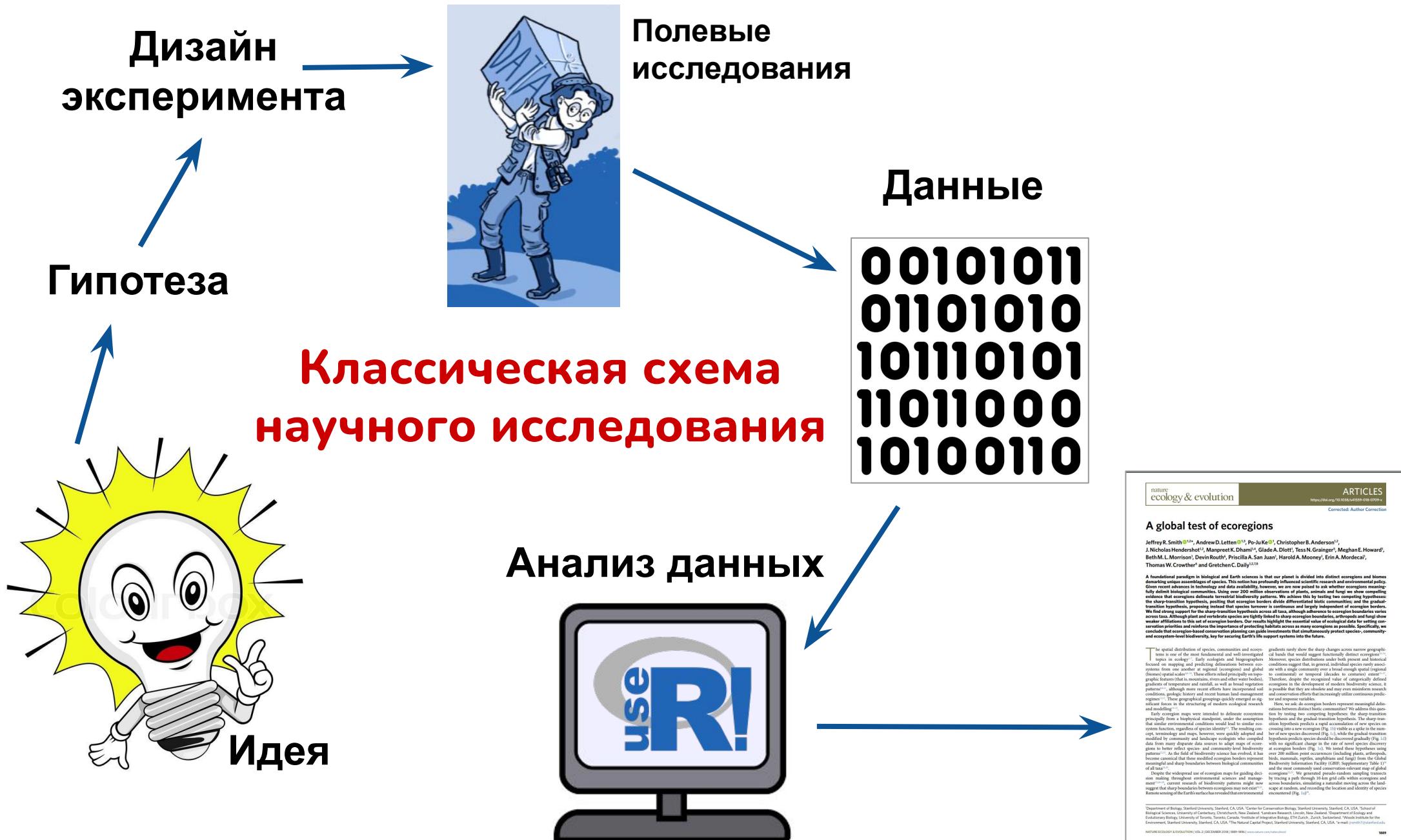


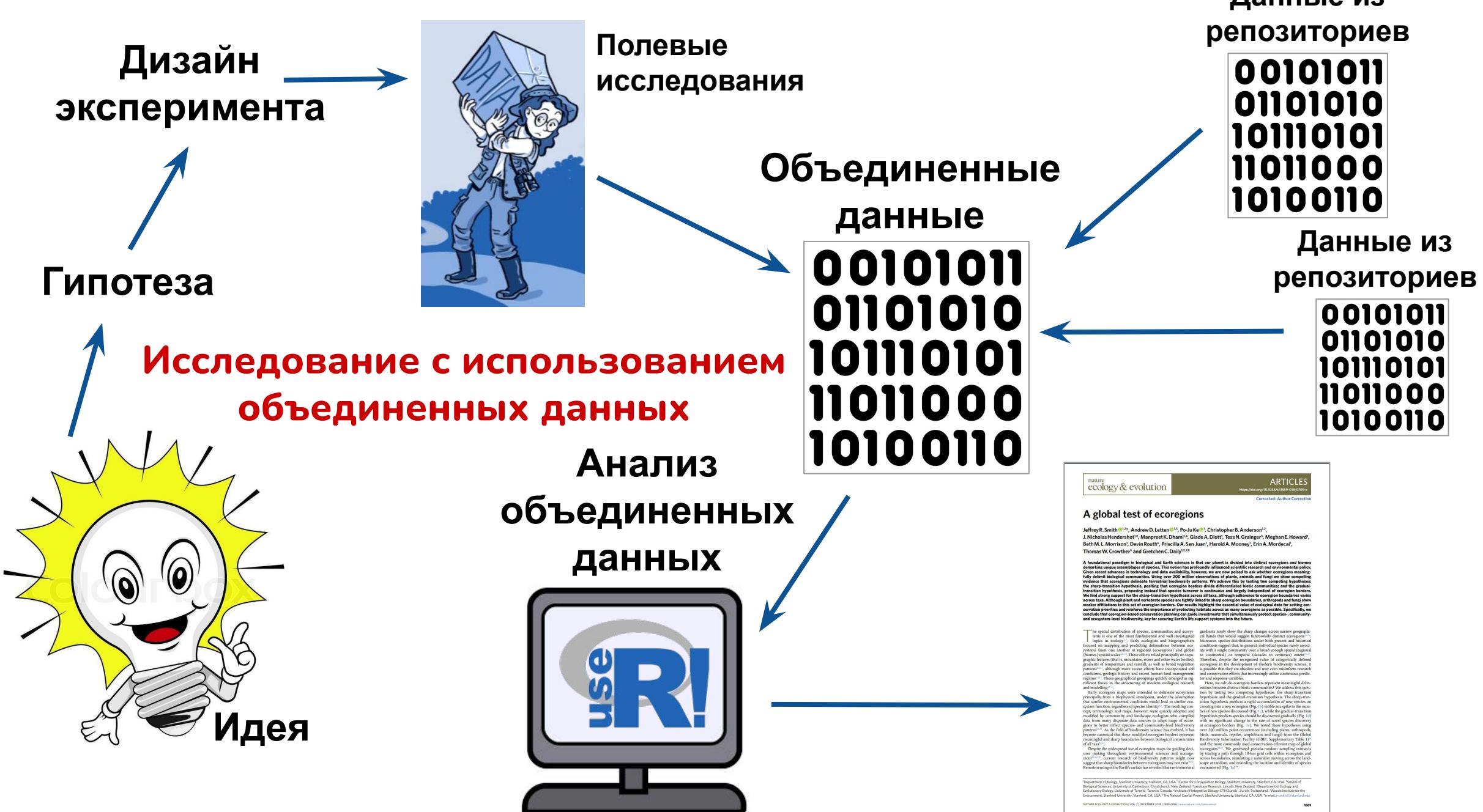
4) ...

**5) PROFIT !!!**

## Программный код - это не так страшно, как кажется







# Использование данных, доступных через GBIF

 NeoBiota

Home Articles About About Pensoft Books

Research Article NeoBiota 35: 61-85  
<https://doi.org/10.3897/neobiota.35.12460> (02 Jun 2017)

## Alien flora of Turkey: checklist, taxonomic composition and ecological attributes

▼ Ahmet Uludag, Necmi Aksoy, Ayşe Yazlık, Zubeyde Filiz Arslan, Efecan Yazmiş, İlhan Antonella Cossu, Quentin Groom, Jan Pergl, Petr Pyšek, Giuseppe Brundu

### Abstract

The paper provides an updated checklist of the alien flora of Turkey with information on its structure. The alien flora of Turkey comprises 340 taxa, among which there are 321 angiosperms, 17 gymnosperms and two ferns. Of the total number of taxa, 228 (68%) are naturalized and 112 (32%) are casual. There are 275 neophytes (172 naturalized and 103 casual) and could not be classified with respect to the res potential to escape are also listed. The richest f and *Solanaceae* (22). As for the naturalized alien (31 taxa), *Poaceae* (22), *Amaranthaceae* (18) a (63.8% of the total number of taxa with this life annuals contribute 33.8% and 2.4% are biennia phanerophytes, of which 20.3% are trees and plants are comparatively less represented. Mos ranges in Americas (44.7%) and Asia (27.6%). 3.8% in Australia and Oceania and 3.5% in the intentionally, whereas the remaining (28.1%) intentionally, the vast majority are ornamental 6.7% as crops. Casual alien plants are most co naturalized taxa are also often recorded (27 naturalized rather than casual (16.0% vs 7.1%, often invaded by alien taxa, especially by those

**Internal Seed Structure of Alpine Plants and Extreme Cold Exposure**

Jaganathan, G. Dalrymple, S. (2019) Data

Cold tolerance in seeds is not well understood compared to mechanisms in aboveground plant tissue but is crucial to understanding how plant populations persist in extreme cold conditions. Counter-intuitively, the ability of seeds to survive extreme cold may become more important in the future due to...

climate change • endosperm • intracellular ice formation • seed mass • snow bed • temperature minima

Journal article | Open access | Peer-reviewed

Data referenced in study DOI 10.15468/14epds DOI 10.15468/3emx9b DOI 10.15468/5sl7sh DOI 10.15468/6e8nje DOI 10.15468/6rl6de DOI 10.15468/7gudy0 DOI 10.15468/a208v DOI 10.15468/aa9jxr DOI 10.15468/ab3s5x DOI 10.15468/avjnkt DOI 10.15468/axrelr DOI 10.15468/axtkuz DOI 10.15468/b5qezo DOI 10.15468/bkzv1l DOI 10.15468/c4w4cc DOI 10.15468/c7c9qa DOI 10.15468/cchsv3 DOI 10.15468/ce7fox DOI 10.15468/cndomv DOI 10.15468/cpnhhc DOI 10.15468/czwtqn DOI 10.15468/djzjge DOI 10.15468/dk5gko DOI 10.15468/dlwwhz DOI 10.15468/dmcdk6 DOI 10.15468/dn6wek DOI 10.15468/eepzxm DOI 10.15468/ekg8cl DOI 10.15468/ewje4c DOI 10.15468/gitoa1 DOI 10.15468/gjluak DOI 10.15468/hja69f DOI 10.15468/hnhr93 DOI 10.15468/hszq1 DOI 10.15468/lawody DOI 10.15468/lb5ypt DOI 10.15468/lrnrwe DOI 10.15468/luxeth DOI 10.15468/jyrthk DOI 10.15468/k64rgi DOI 10.15468/kasmwk DOI 10.15468/kcyxvg DOI 10.15468/kiosdm DOI 10.15468/kkpr3 DOI 10.15468/kowta4 DOI 10.15468/kvfygp DOI 10.15468/llipc DOI 10.15468/ly60bx DOI 10.15468/mpsjrk DOI 10.15468/mypdjd

**Complex agroforestry systems against biotic homogenization: The case of plants in the herbaceous stratum of cocoa production systems**

Fridley JD, Peet RK, Wentworth TR, White PS. 2005. Connecting fine- and broad-scale species-area relationships of southeastern U.S. flora. *Ecology* 86: 1172–1177.

Friedman JH. 2002. Stochastic gradient boosting. *Computational Statistics & Data Analysis* 38: 367–378.

Freeman S, Garrouette MD. 2012. Vascular plants of Buldir Island, Aleutian Islands, Alaska. Unpublished.

[GBIF Occurrence Download](http://doi.org/10.15468/dl.0sude6) <http://doi.org/10.15468/dl.0sude6>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.2d1n1j) <http://doi.org/10.15468/dl.2d1n1j>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.2ep9fe) <http://doi.org/10.15468/dl.2ep9fe>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.3e1j6) <http://doi.org/10.15468/dl.3e1j6>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.3iydrb) <http://doi.org/10.15468/dl.3iydrb>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.3l2mn) <http://doi.org/10.15468/dl.3l2mn>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.3xp6yq) <http://doi.org/10.15468/dl.3xp6yq>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.4dgula) <http://doi.org/10.15468/dl.4dgula>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.4oehhl) <http://doi.org/10.15468/dl.4oehhl>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.5e0lqo) <http://doi.org/10.15468/dl.5e0lqo>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.7l2gk) <http://doi.org/10.15468/dl.7l2gk>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.8qm4rk) <http://doi.org/10.15468/dl.8qm4rk>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.9nnht) <http://doi.org/10.15468/dl.9nnht>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.9t0izv) <http://doi.org/10.15468/dl.9t0izv>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.9ymnm0) <http://doi.org/10.15468/dl.9ymnm0>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.aataa8) <http://doi.org/10.15468/dl.aataa8>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.ajqbio) <http://doi.org/10.15468/dl.ajqbio>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.aoahf8b) <http://doi.org/10.15468/dl.aoahf8b>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.ay9irt) <http://doi.org/10.15468/dl.ay9irt>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.bge3hq) <http://doi.org/10.15468/dl.bge3hq>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.bjxpd1p) <http://doi.org/10.15468/dl.bjxpd1p>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.bn3x3e) <http://doi.org/10.15468/dl.bn3x3e>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.cwvjsd) <http://doi.org/10.15468/dl.cwvjsd>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.cyacpq) <http://doi.org/10.15468/dl.cyacpq>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.dep98) <http://doi.org/10.15468/dl.dep98>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.dflhagu) <http://doi.org/10.15468/dl.dflhagu>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.dl7p19) <http://doi.org/10.15468/dl.dl7p19>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.diyll) <http://doi.org/10.15468/dl.diyll>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.e3zjja) <http://doi.org/10.15468/dl.e3zjja>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.evd6zp) <http://doi.org/10.15468/dl.evd6zp>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.f3t0h4) <http://doi.org/10.15468/dl.f3t0h4>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.f4nfra) <http://doi.org/10.15468/dl.f4nfra>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.fon17x) <http://doi.org/10.15468/dl.fon17x>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.gdkhoff) <http://doi.org/10.15468/dl.gdkhoff>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.gfkdx) <http://doi.org/10.15468/dl.gfkdx>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.h7ofzk) <http://doi.org/10.15468/dl.h7ofzk>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.hrfqql) <http://doi.org/10.15468/dl.hrfqql>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.hrqyvo) <http://doi.org/10.15468/dl.hrqyvo>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.idisdq) <http://doi.org/10.15468/dl.idisdq>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.ihefj) <http://doi.org/10.15468/dl.ihefj>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.iihmop) <http://doi.org/10.15468/dl.iihmop>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.ilzhkw) <http://doi.org/10.15468/dl.ilzhkw>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.jqlclwx) <http://doi.org/10.15468/dl.jqlclwx>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.jthdg0) <http://doi.org/10.15468/dl.jthdg0>

[GBIF Occurrence Download](http://doi.org/10.15468/dl.kbewa9) <http://doi.org/10.15468/dl.kbewa9>

**nature ecology & evolution**

ARTICLES

<https://doi.org/10.1038/s41559-018-0709-x>

Corrected: Author Correction

## A global test of ecoregions

Jeffrey R. Smith<sup>1,2\*</sup>, Andrew D. Letten<sup>1,3</sup>, Po-Ju Ke<sup>1</sup>, Christopher B. Anderson<sup>1,2</sup>, J. Nicholas Hendershot<sup>1,4</sup>, Manpreet K. Dhami<sup>1,4</sup>, Glade A. Dlott<sup>1</sup>, Tess N. Grainger<sup>5</sup>, Meghan E. Howard<sup>1</sup>, Beth M. L. Morrison<sup>1</sup>, Devin Routh<sup>6</sup>, Priscilla A. San Juan<sup>1</sup>, Harold A. Mooney<sup>1</sup>, Erin A. Mordecai<sup>1</sup>, Thomas W. Crowther<sup>4</sup> and Gretchen C. Daily<sup>1,2,7,8</sup>

A foundational paradigm in biological and Earth sciences is that our planet is divided into distinct ecoregions and biomes, marking unique assemblages of species. This notion has profoundly influenced scientific research and environmental policy. Given recent advances in technology and data availability, however, we are now poised to ask whether ecoregions meaningfully delimit biological communities. Using over 200 million observations of plants, animals and fungi we show compelling evidence that ecoregions delineate terrestrial biodiversity patterns. We achieve this by testing two competing hypotheses: the sharp-transition hypothesis, positing that ecoregion borders divide differentiated biotic communities; and the gradual-transition hypothesis, proposing instead that species turnover is continuous and largely independent of ecoregion borders. We find strong support for the sharp-transition hypothesis across all taxa, although adherence to ecoregion boundaries varies across taxa. A single ecoregion's boundaries are tight, linking sharp ecoregion borders, arthropods and fungi show weak associations to this set of ecoregion borders. Our results highlight the central value of ecological data for setting conservation priorities and reinforce the importance of protecting habitats across as many ecoregions as possible. Specifically, we conclude that ecoregion-based conservation planning can guide investments that simultaneously protect species, community and ecosystem-level biodiversity, key for securing Earth's life support systems into the future.

The spatial distribution of species, communities and ecoregions is one of the most fundamental and well-investigated topics in ecology. Previous studies have focused on mapping and predicting delineations between ecosystems from one another at regional (ecoregions) and global (biomes) spatial scales<sup>1–3</sup>. These efforts relied principally on topographic features (that is, mountains, rivers and other water bodies), gradients of temperature and rainfall, as well as broad vegetation patterns<sup>4,5</sup>, although more recent efforts have incorporated soil conditions, geologic history and recent human land-management regimes<sup>1,2,6</sup>. These geographical groupings quickly emerged as significant features in the structuring of modern ecological research and modelling<sup>1,2,7</sup>.

Early ecoregion maps were intended to delineate ecosystems principally from a biological standpoint. The assumption was that similar environmental conditions would lead to similar ecosystem function, regardless of species identity<sup>8</sup>. The resulting concept, terminology and maps, however, were quickly adopted and modified by community and landscape ecologists who compiled data from many disparate data sources to adapt maps of ecoregions to better reflect species- and community-level biodiversity patterns<sup>1,2,3</sup>. As the field of biodiversity science has evolved, it has become canonical that these modified ecoregion borders represent meaningful and sharp boundaries between biological communities of all taxa<sup>1,2</sup>.

Despite the widespread use of ecoregion maps for guiding decision making throughout environmental sciences and management<sup>1,2,3,4,5</sup>, current research of biodiversity patterns might now suggest that sharp boundaries between ecoregions do not exist<sup>1,2</sup>. We generated pseudo-random sampling transects by tracing a path through 10-km grid cells within ecoregions and across boundaries, simulating a naturalist moving across the landscape at random, and recording the location and identity of species encountered (Fig. 1a).

<sup>1</sup>Department of Biology, Stanford University, Stanford, CA, USA. <sup>2</sup>Center for Conservation Biology, Stanford University, Stanford, CA, USA. <sup>3</sup>School of Biological Sciences, University of Canterbury, Christchurch, New Zealand. <sup>4</sup>Landcare Research, Lincoln, New Zealand. <sup>5</sup>Department of Ecology and Evolutionary Biology, University of Toronto, Toronto, Canada. <sup>6</sup>Institute of Integrative Biology, ETH Zurich, Zurich, Switzerland. <sup>7</sup>Woods Institute for the Environment, Stanford University, Stanford, CA, USA. <sup>8</sup>The Natural Capital Project, Stanford University, Stanford, CA, USA. \*e-mail: jsmith7@stanford.edu

NATURE ECOLOGY & EVOLUTION | VOL 2 | DECEMBER 2018 | 1889–1896 | www.nature.com/natecol

1889

- Kenney LA, Kanner R. 2010. Vascular plants of Agattu Island, Aleutian Islands, Alaska. Unpublished.
- Klein DR. 1968. Introduction, increase, and crash of reindeer on St. Matthew Island. *Journal of Wildlife Management* 32: 350–367.
- Kreft H, Jetz W. 2007. Global patterns and determinants of vascular plant diversity. *Proceedings of the National Academy of Sciences USA* 104: 5925–5930.

# Цитирование данных: отслеживание и отображение

OCCURRENCE DATASET | REGISTERED OCTOBER 13, 2016

## NMNH Extant Specimen Records

Published by [National Museum of Natural History, Smithsonian Institution](#)

Thomas Orrell

7,571,188 OCCURRENCES

218 CITATIONS

Read more about literature, how it's discovered and linked to GBIF-mediated data.

**Leptodactylus mystaceus (Spix, 1824): Distribution extension for the Brazilian Cerrado (Anura: Leptodactylidae)**

Andreani, T. Borges, R. Santos, L. (2017) Herpetology Notes

Fifty-seven species of anurans belonging to the genus *Leptodactylus* are registered and widely distributed throughout Brazil (Sá et al., 2014; SBH, 2016). Among these species, *Leptodactylus mystaceus* (Spix, 1824) spreads throughout the whole Northern region and some areas of northeastern and central ...

Journal Article

Data used in study DOI [10.15468/dl.s1xtf6](https://doi.org/10.15468/dl.s1xtf6)

**DO ENVIRONMENTAL FACTORS AFFECT THE TAXONOMIC RELIABILITY OF LEAF CUTICULAR MICROMORPHOLOGICAL CHARACTERS? A CASE STUDY IN PODOCARPACEAE**

Clugston, J. Jeffree, C. Ahrends, A. Mill, R. (2017) Edinburgh Journal of Botany

Leaf cuticle micromorphology has been cited as an important set of taxonomic characters in gymnosperms, but previous studies have largely been based on small sample sizes. The premise of this study is to understand whether external factors affect cuticular micromorphology of Podocarpaceae. Two exam...

Journal Article

Data used in study DOI [10.15468/6e8nje](https://doi.org/10.15468/6e8nje) DOI [10.15468/bkzv1l](https://doi.org/10.15468/bkzv1l) DOI [10.15468/dgbpla](https://doi.org/10.15468/dgbpla) DOI [10.15468/dlvu12](https://doi.org/10.15468/dlvu12) DOI [10.15468/hja69f](https://doi.org/10.15468/hja69f) DOI [10.15468/hnhrg3](https://doi.org/10.15468/hnhrg3) DOI [10.15468/i9bj5r](https://doi.org/10.15468/i9bj5r) DOI [10.15468/ib5ypt](https://doi.org/10.15468/ib5ypt) DOI [10.15468/ly60bx](https://doi.org/10.15468/ly60bx) DOI [10.15468/mug7kr](https://doi.org/10.15468/mug7kr) DOI [10.15468/nc6rxv](https://doi.org/10.15468/nc6rxv) DOI [10.15468/ucmdjy](https://doi.org/10.15468/ucmdjy) DOI [10.15468/vtfbe3](https://doi.org/10.15468/vtfbe3) DOI [10.15468/x5ucvh](https://doi.org/10.15468/x5ucvh) DOI [10.15468/xy0eo1](https://doi.org/10.15468/xy0eo1) DOI [10.15468/yo3mmu](https://doi.org/10.15468/yo3mmu) DOI [10.15468/ypoair](https://doi.org/10.15468/ypoair) DOI [10.15468/z7ps7](https://doi.org/10.15468/z7ps7)

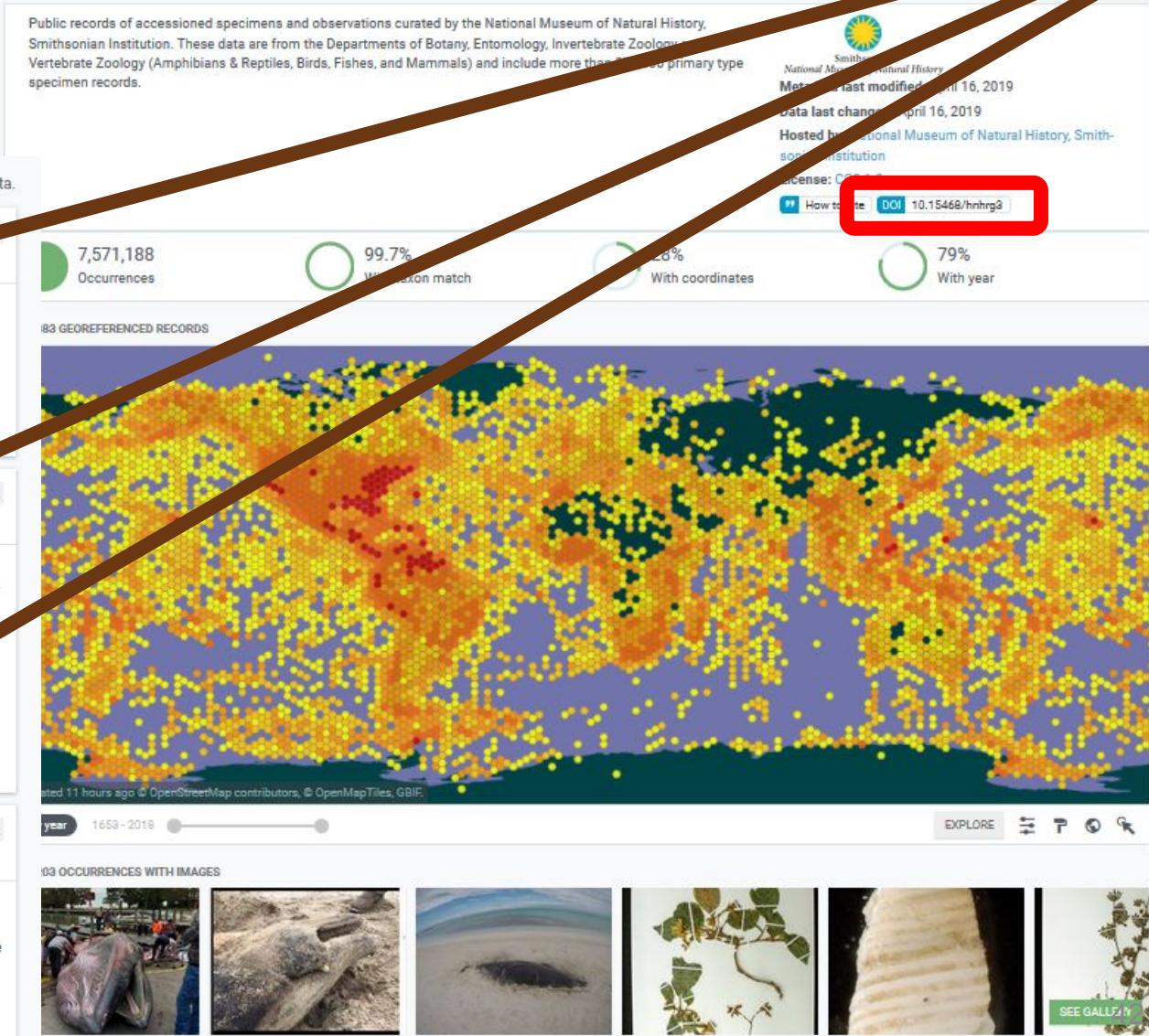
**One for each ocean: revision of the *Bursa granularis* (Röding, 1798) species complex (Gastropoda: Tonnaidea: Bursidae)**

Sanders, M. Merle, D. Bouchet, P. Castelnau, M. Beu, A. Samadi, S. - (2017) Journal of Molluscan Studies

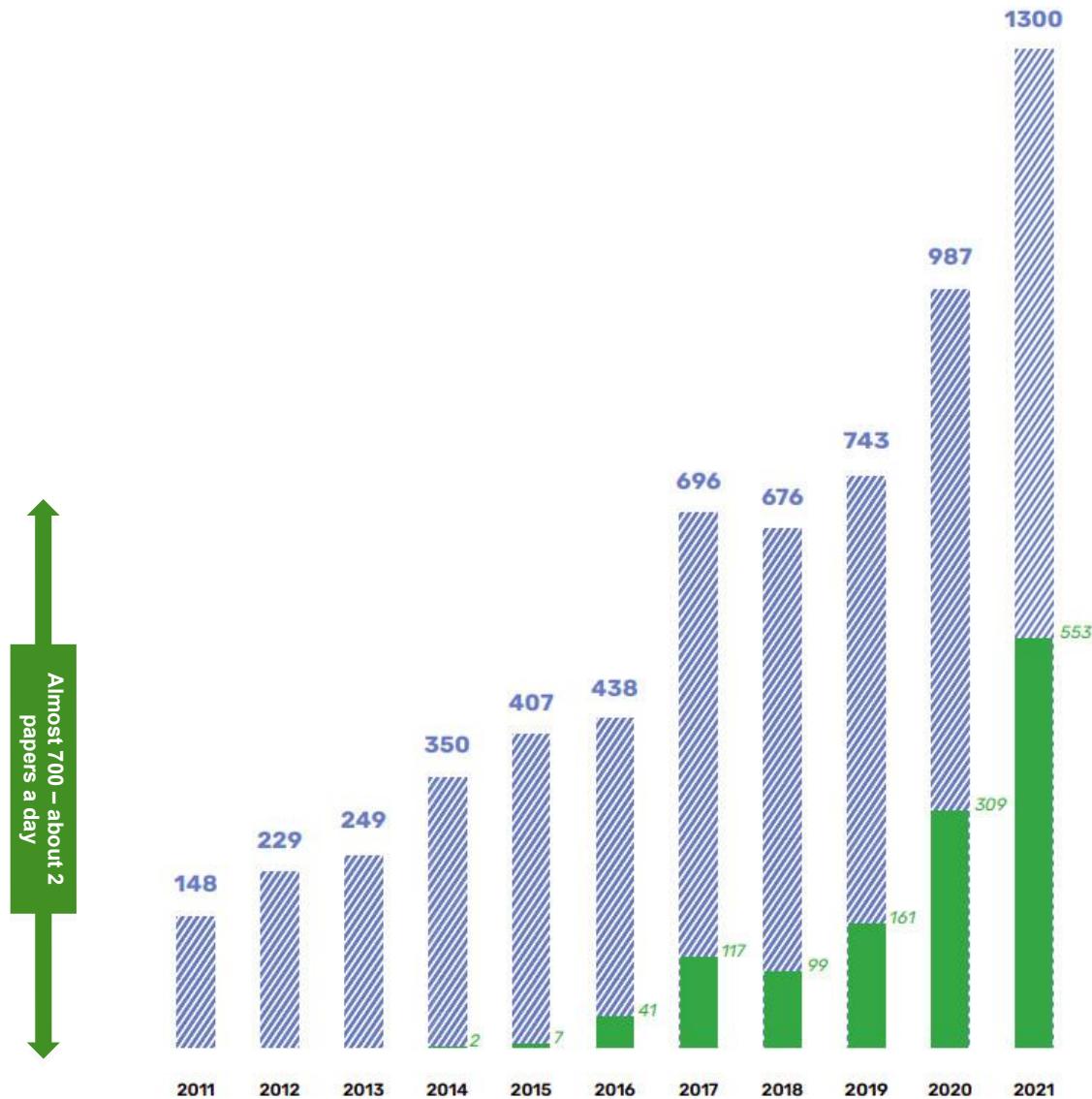
*Bursa granularis* (Röding, 1798) is a tonnaeid gastropod that is regarded as broadly distributed throughout the Indo-Pacific and tropical western Atlantic. Because of its variable shell, it has received no less than thirteen names, now all synonymized under the name *B. granularis*. We sequenced a fra...

Journal Article

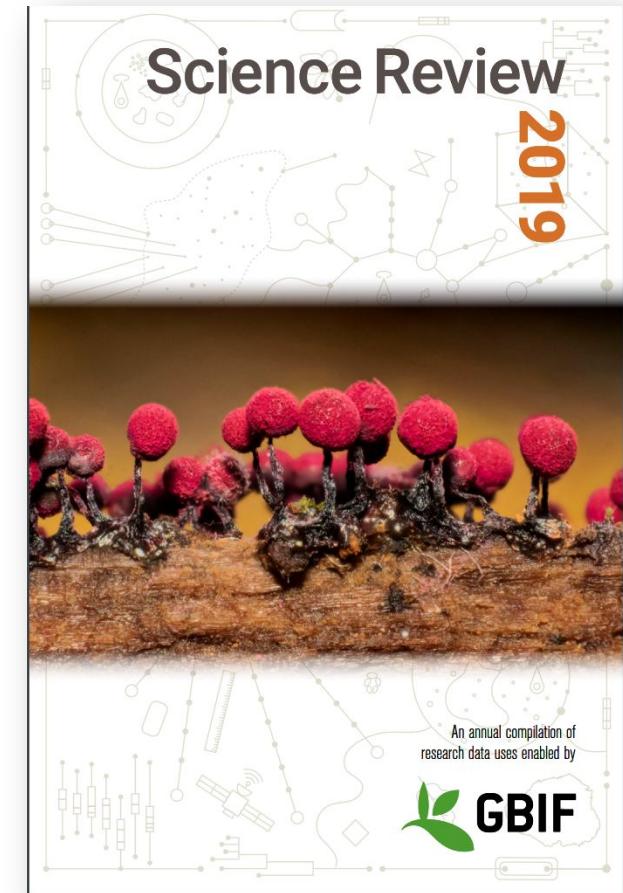
Data used in study DOI [10.15468/dl.ma8d1e](https://doi.org/10.15468/dl.ma8d1e)



# Научные публикации, использующие данные GBIF

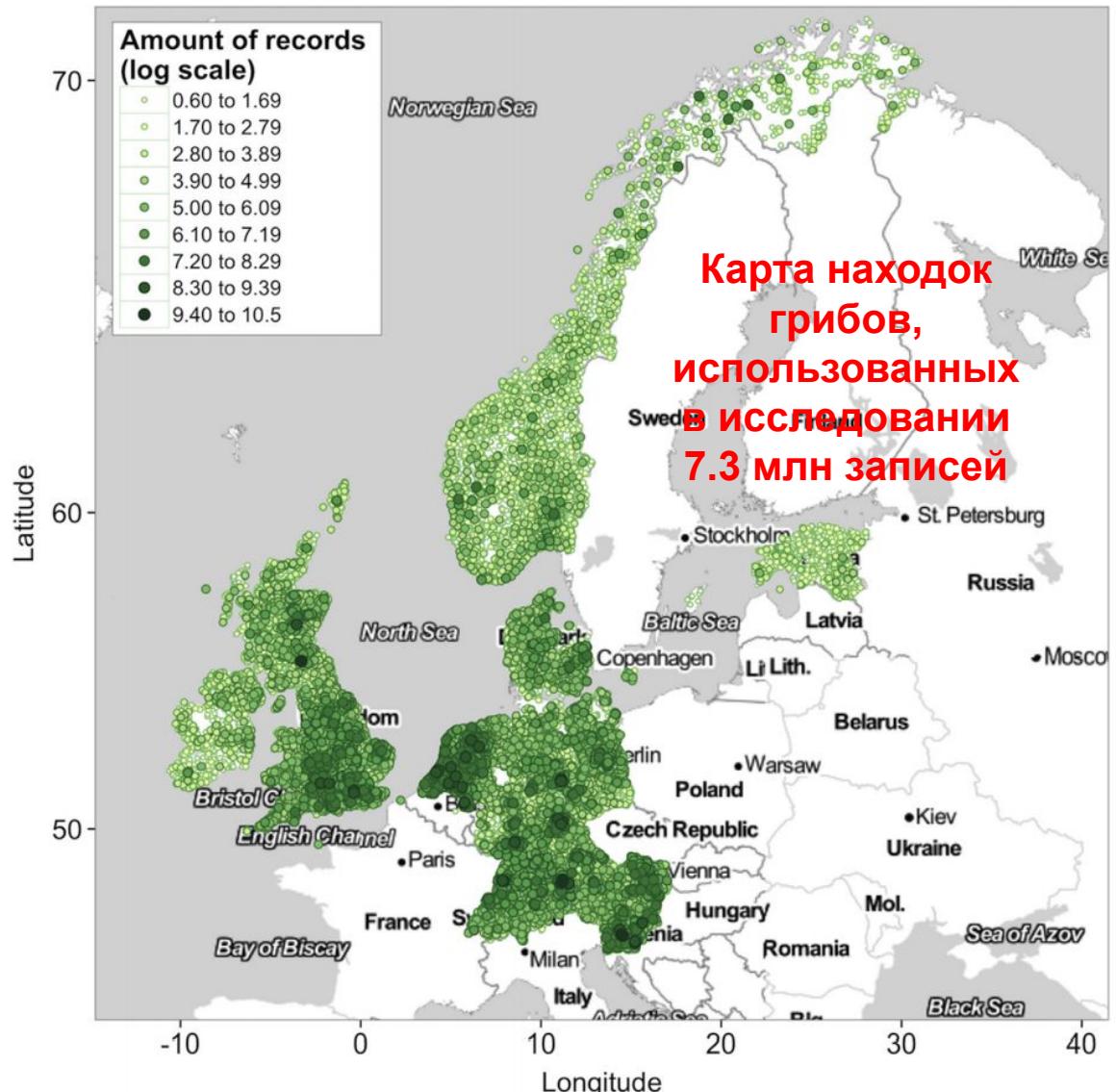


<https://www.gbif.org/science-review>  
[https://www.gbif.org/sites/default/files/gbif\\_analytics/country/RU/GBIF\\_CountryReport\\_RU.pdf](https://www.gbif.org/sites/default/files/gbif_analytics/country/RU/GBIF_CountryReport_RU.pdf)



# Big data integration: Pan-European fungal species observations assembly for addressing contemporary questions in ecology and global change biology

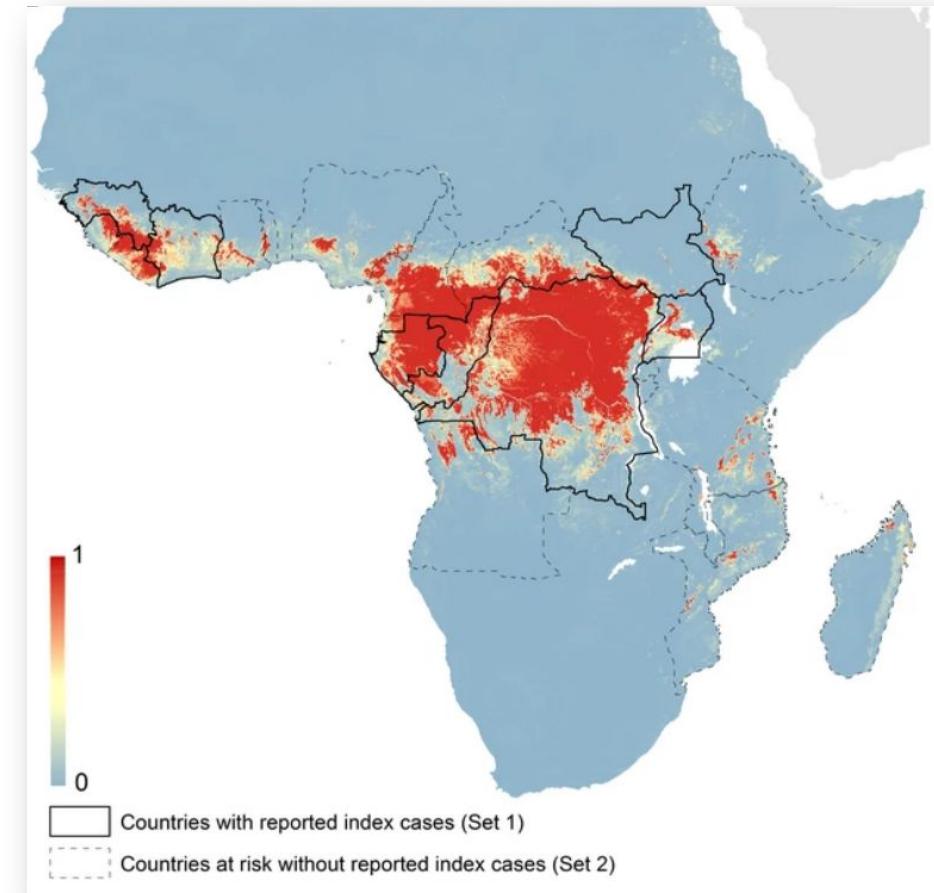
- Таксономический анализ
- Анализ сходства сообществ грибов в разных регионах
- Фенология появления плодовых тел
- Научный потенциал собранного набора данных
- Анализ данных с позиций глобальных изменений климата



# Mapping the zoonotic niche of Ebola virus disease in Africa

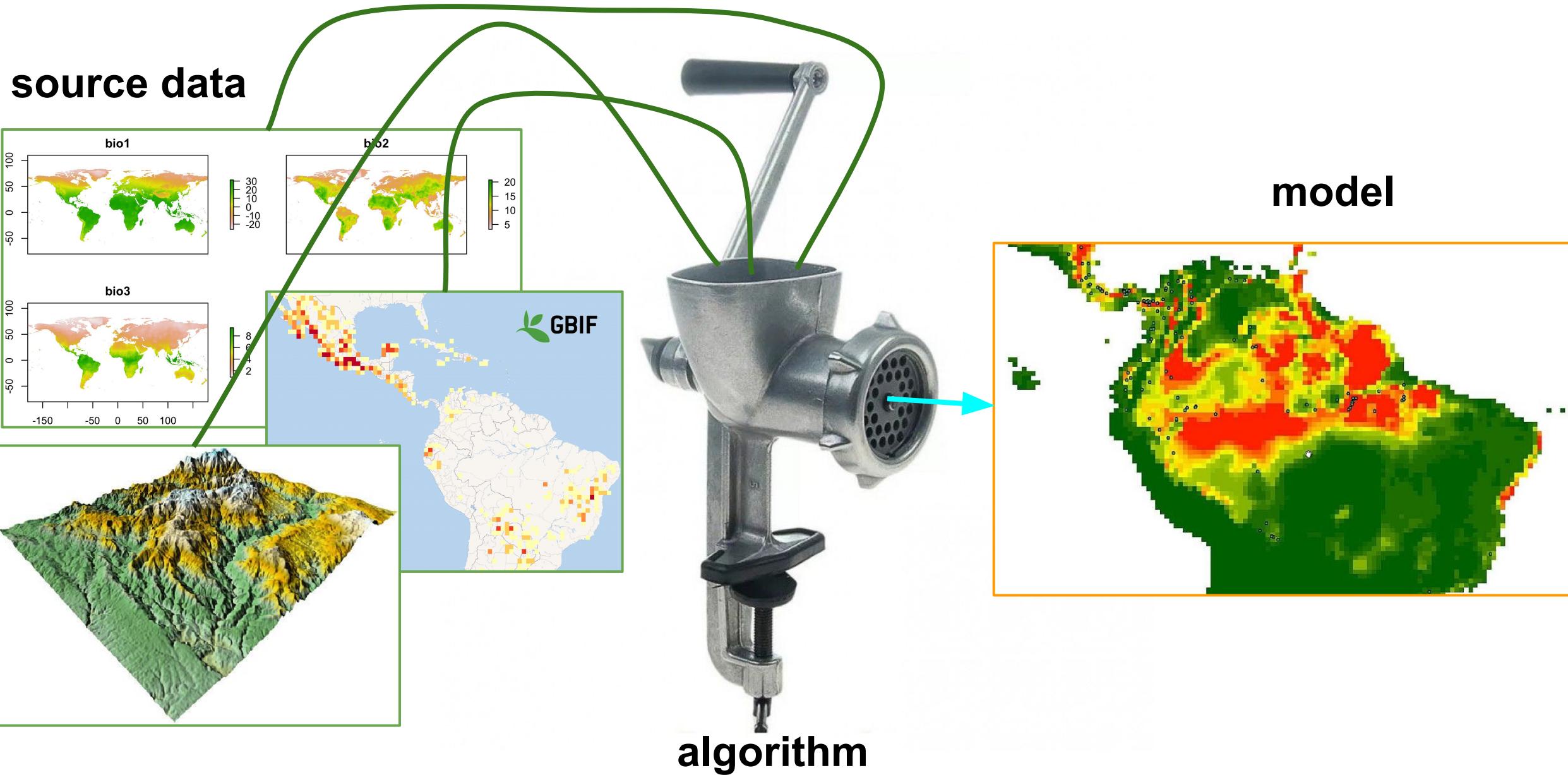


- Смоделированы экологические ниши трех видов летучих мышей, связанных с передачей вируса Эбола
- Данные получены из глобальной базы по биоразнообразию GBIF
- Результаты показали, что зоны риска охватывают 22 страны, население 22 миллиона чел.
- Результаты помогают расставить приоритеты в области эпидем. надзора и диагностики в зонах повышенного риска



Распространение вируса Эбола, предсказанное моделью

# Species Distribution Modelling (SDM)



# Глобальное распределение разнообразия дождевых червей

6928 точек сбора в 57 странах

Three generalized linear mixed-effects models

- A
1. Видовое богатство
  2. Численность
  3. Биомасса

DOI: [10.1126 /science.aax4851](https://doi.org/10.1126/science.aax4851)

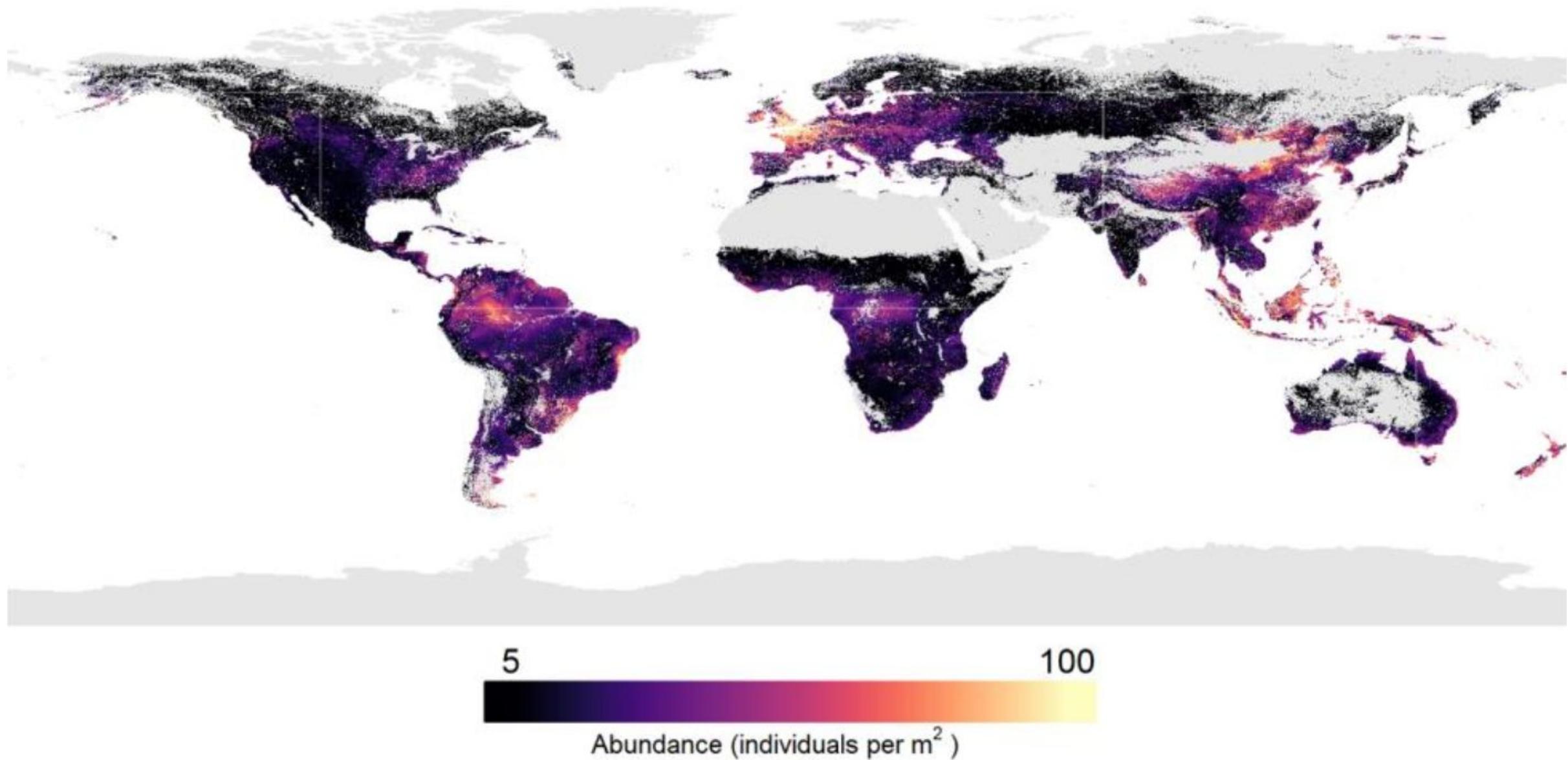
Number Of Sites

- 1
- 100
- 200
- 300
- 400
- 500
- 600

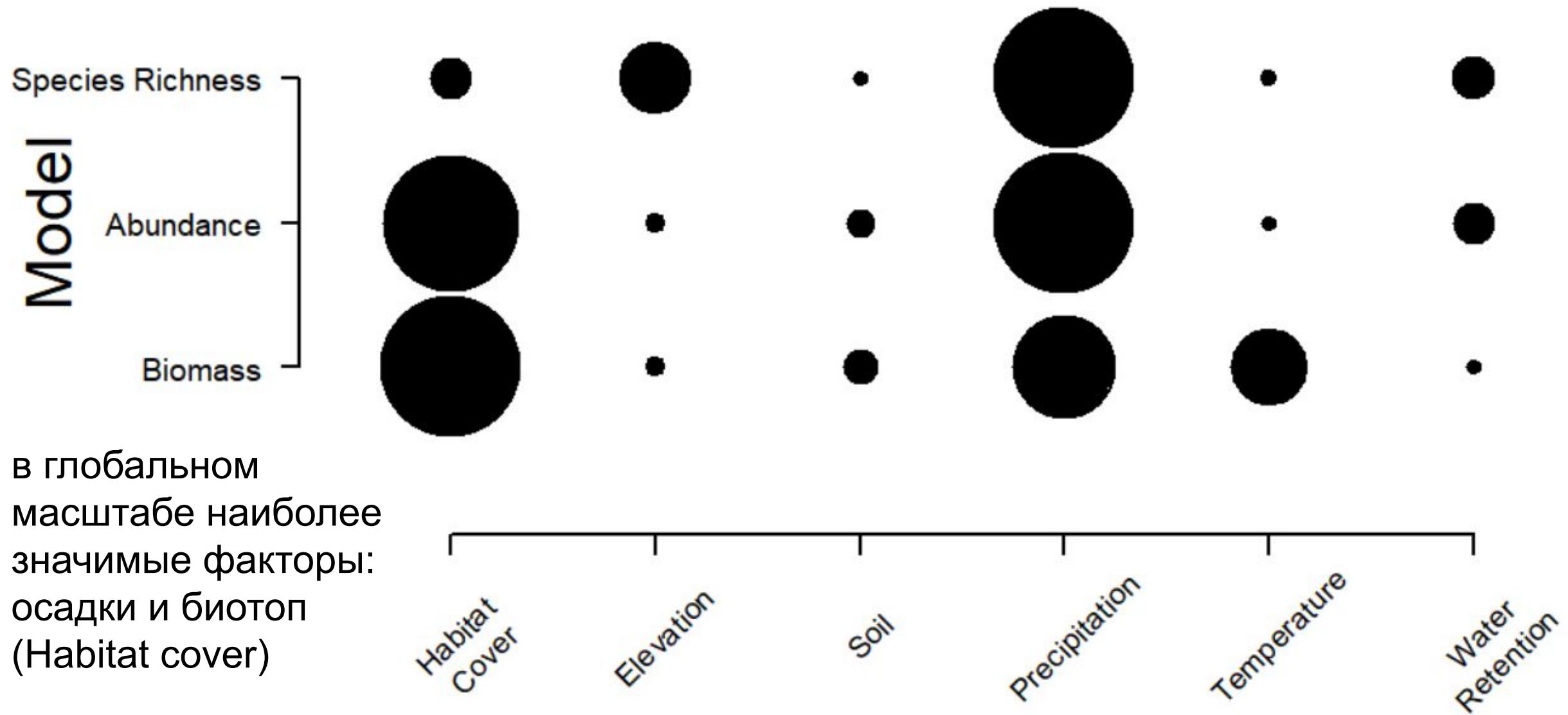
Предикторы:  
Habitat cover  
Elevation  
Soil  
Precipitation  
Temperature  
Water retention



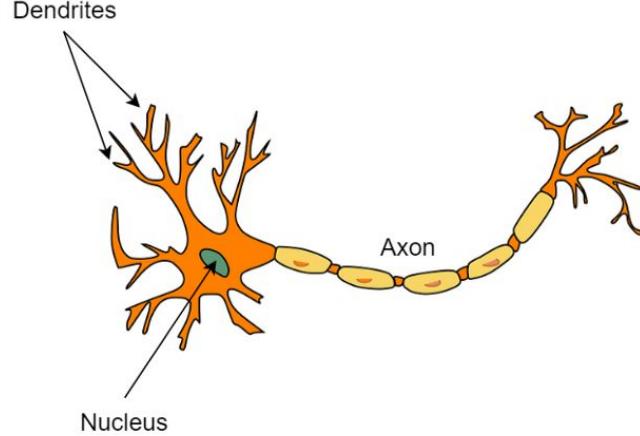
# Распределение обилия, в экз. на кв. м



# Влияние факторов на население дождевых червей



# Искусственные нейронные сети

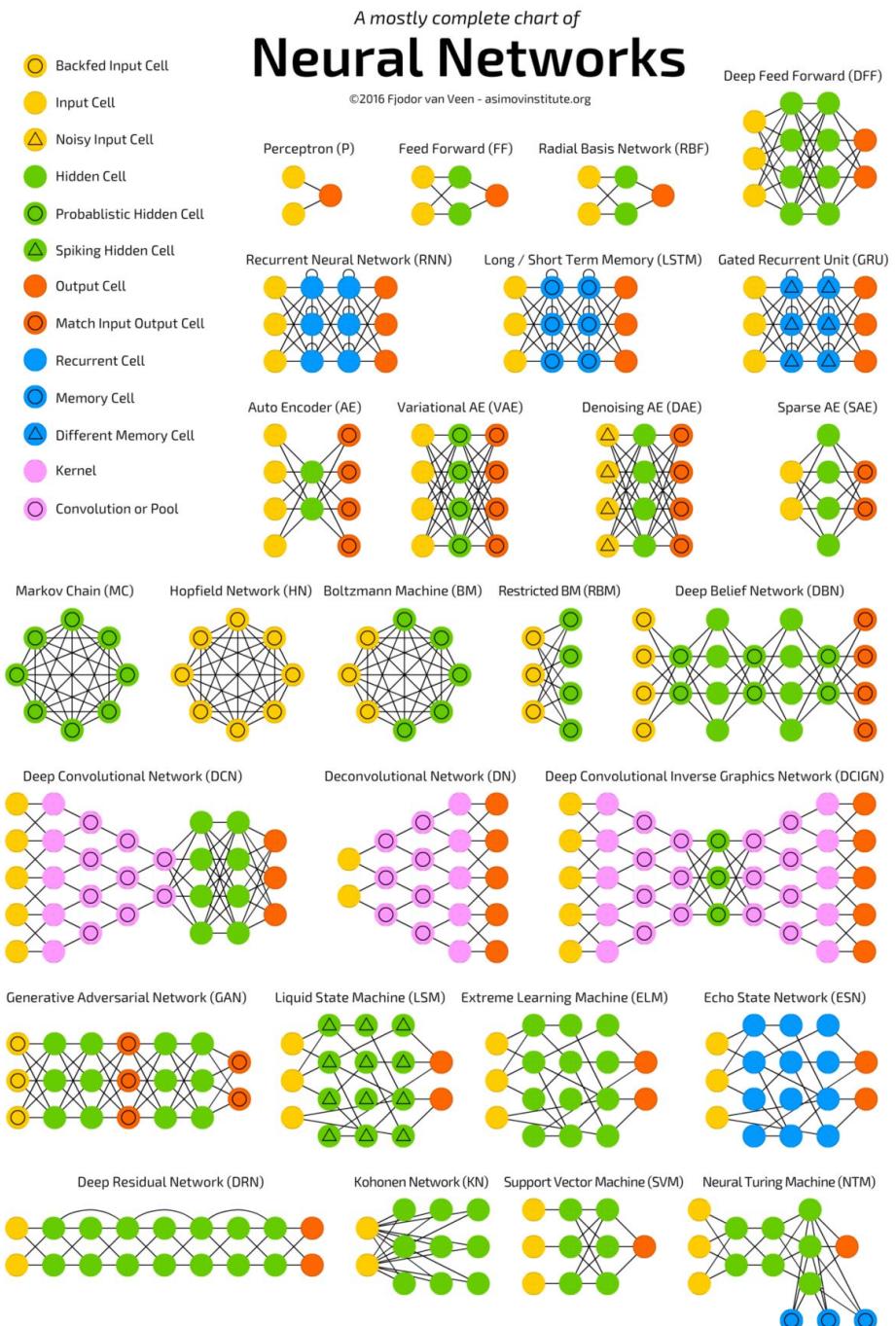


Для выполнения задачи нет заранее определенного алгоритма, но есть архитектура нейронной сети разработанная для решения какого-то определенного класса задач

У каждого искусственного нейрона есть свой вес - "мощность исходящего сигнала"

Обучение нейросети - стохастический процесс, который меняет вес нейронов

В процессе обучения происходит кросс-валидация на тестовой выборке размеченных данных



# Сверточные нейронные сети

1 $x_1$	1 $x_0$	1 $x_1$	0	0
0 $x_0$	1 $x_1$	1 $x_0$	1	0
0 $x_1$	0 $x_0$	1 $x_1$	1	1
0	0	1	1	0
0	1	1	0	0

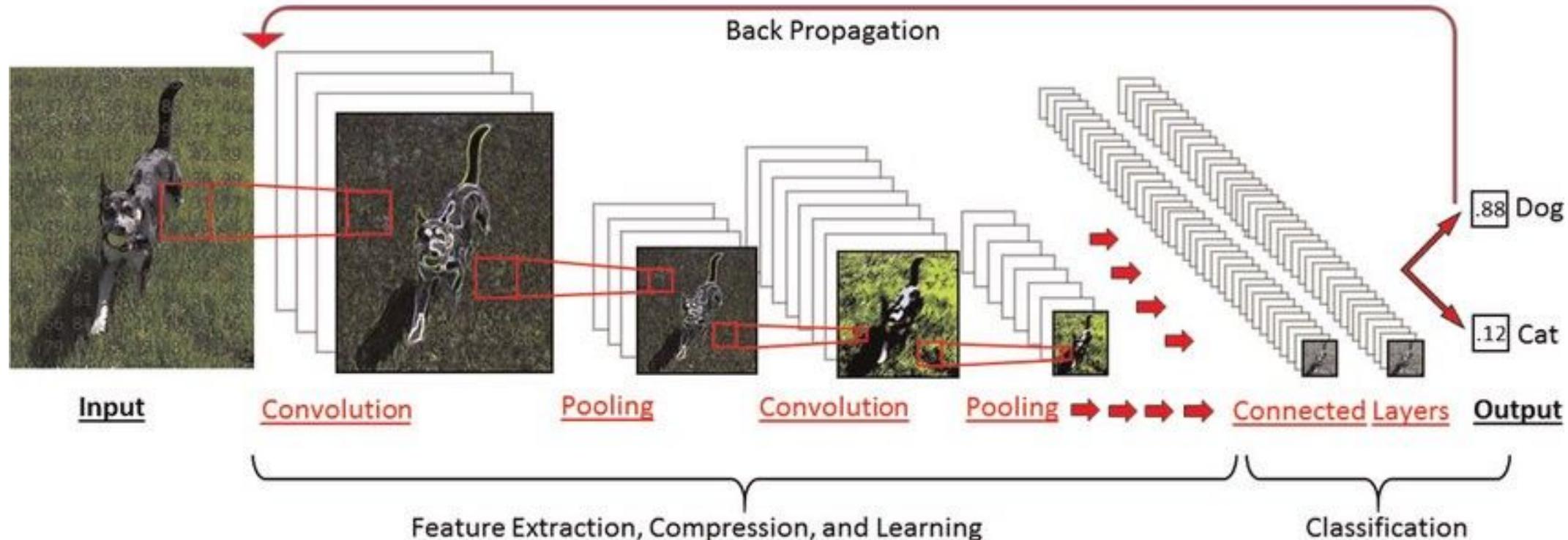
Image

4		

Convolved  
Feature

Convolutional Neural Network  
основа работы сети - свёртка изображений

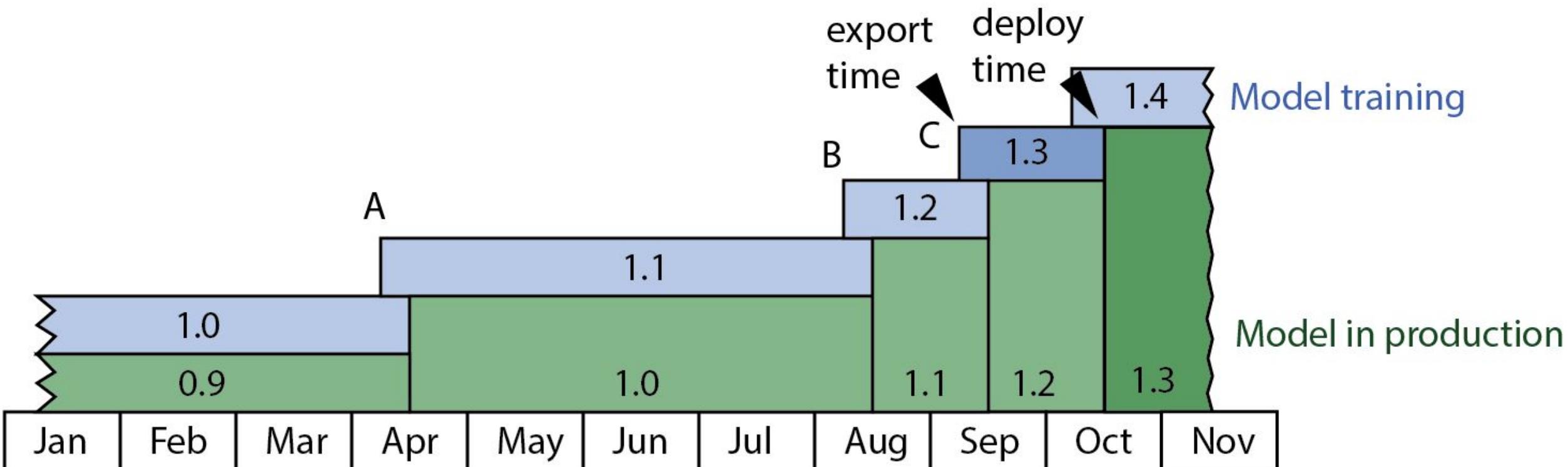
## Deep Learning with Convolutional Neural Networks

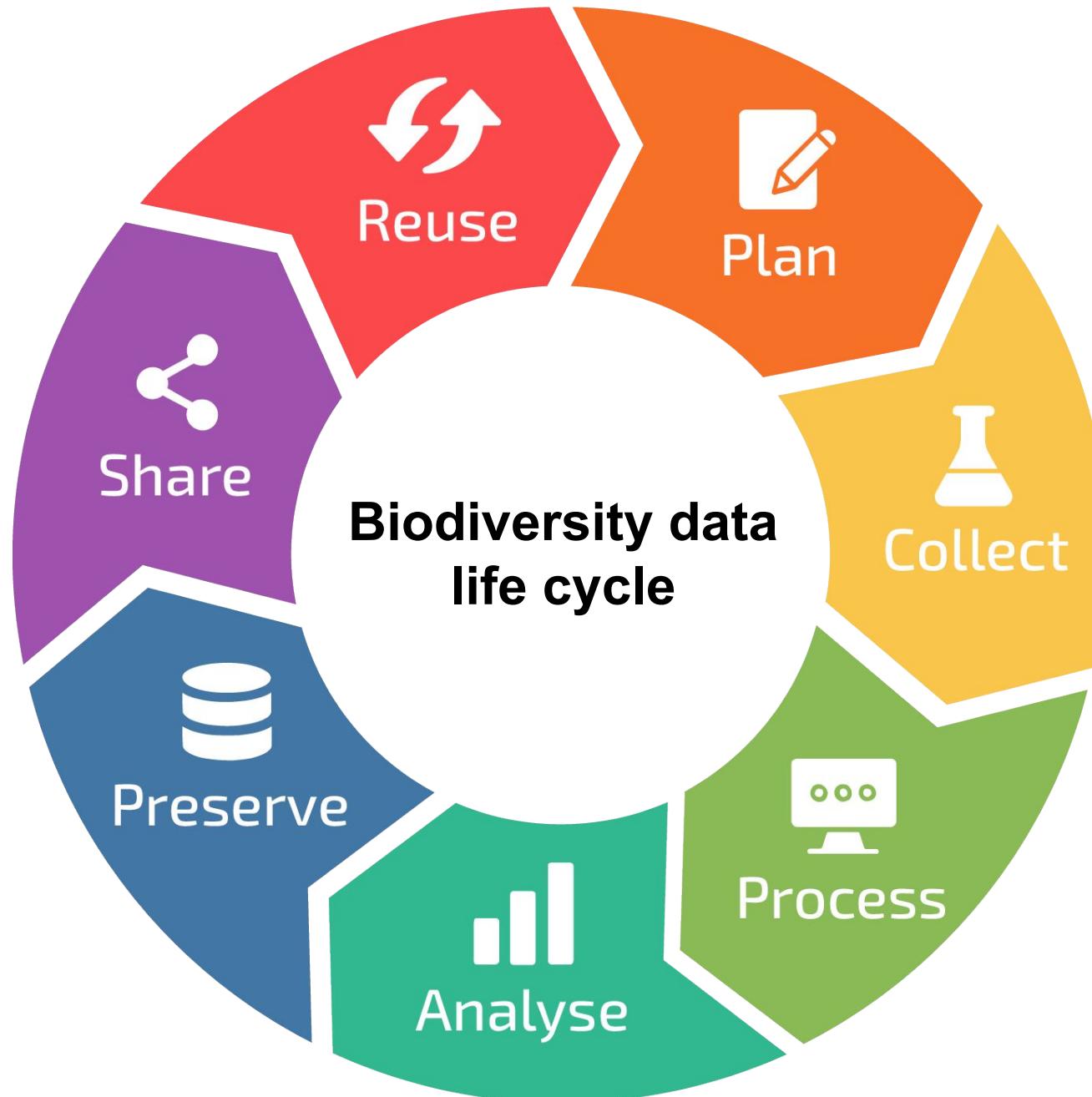


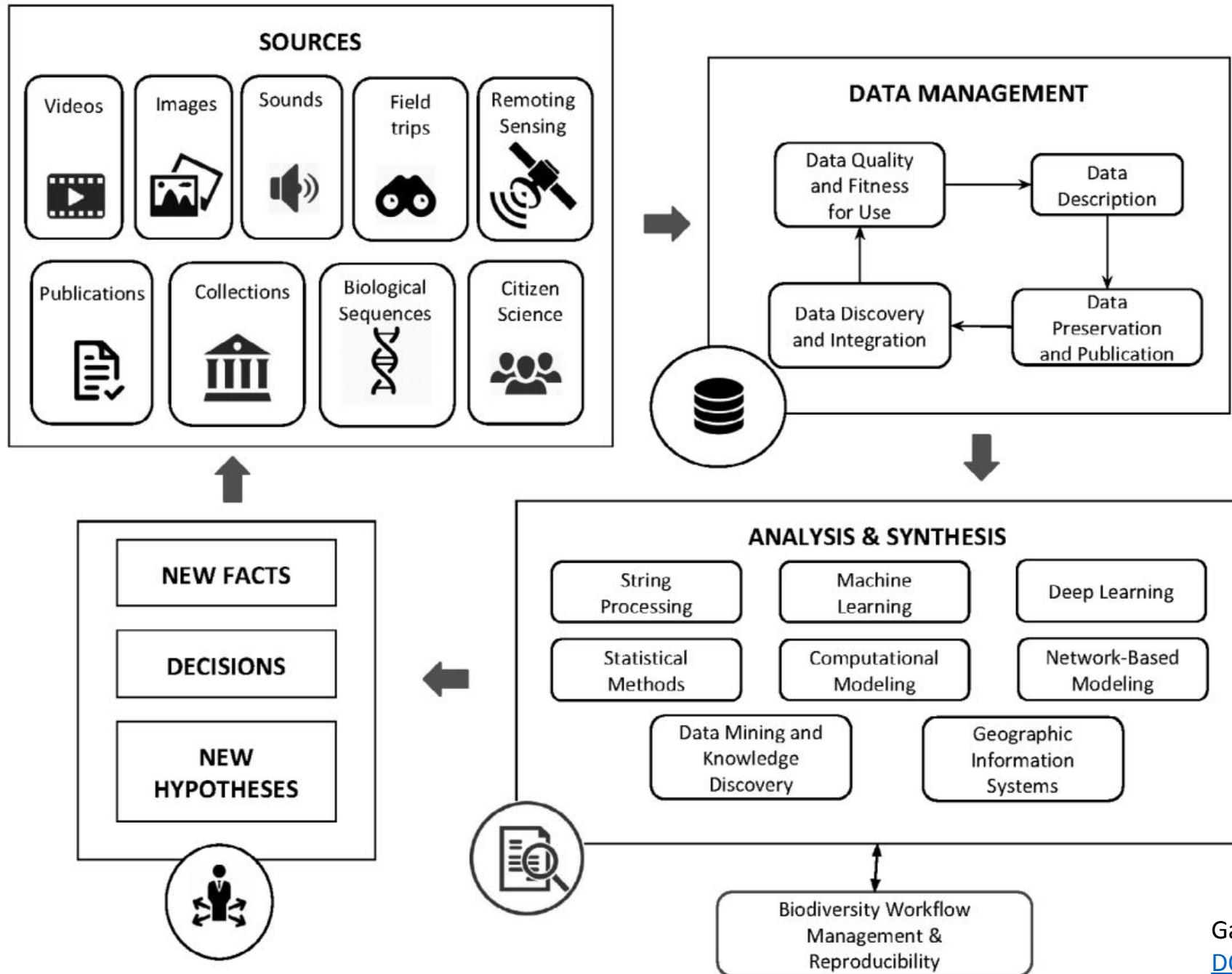
# Нейронная сеть в iNaturalist

первая система распознавания изображений была запущена в 2017 году, включала 5089 таксонов. Обучение проводилось на 579 184 учебных изображений и 95 986 тестовых.

Актуальная Computer Vision Model включает 66 214 таксонов, для каждого из которых отобрано как минимум 50 качественных изображений







# International Biodiversity Informatics Landscape

## Информатика биоразнообразия в мире

### Vocabularies & Standards

Biodiversity Information Standards  
TDWG



Catalogue of Life

### International Portals



**GBIF**

Global Biodiversity Information Facility



**BOLD SYSTEMS**



**OBIS**  
OCEAN BIOGEOGRAPHIC INFORMATION SYSTEM

### Soft & Data analysis



### Citizen science

**eBird**



### Digitization



### Share



# Материалы для самостоятельного изучения

Luiz M. R. Gadelha Jr et al., 2020. [A survey of biodiversity informatics: Concepts, practices, and challenges](#)

Bisby F.A. (2000) The quiet revolution: Biodiversity informatics and the internet. *Science*, 289: 2309-2312  
DOI: [10.1126/science.289.5488.2309](https://doi.org/10.1126/science.289.5488.2309)

Edwards J.L., Lane M.L, and Nielsen E.S. (2000) Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop. *Science*, 289, 5488, 2312-2314 DOI: [10.1126/science.289.5488.2312](https://doi.org/10.1126/science.289.5488.2312)

**The future is already here —  
it's just not very evenly distributed**

*Уильям Форд Гибсон,  
писатель-фантаст*

