Естественнонаучные коллекции Югры: сбор, фиксация, хранение, введение в научный оборот
Ханты-Мансийск, 30 марта — 1 апреля 2021

Использование данных GBIF в научных исследованиях: введение

Наталья Иванова[1]
Менторы: **Николай Груданов[2], Нина Филиппова[3], Иван Чадин[4], Денис Сандanов[5]**

1 – ИМПБ РАН - филиал ИПМ им. М.В. Келдыша РАН (Пущино)
2 – УрФУ (Екатеринбург)
3 – ЮГУ (Ханты-Мансийск)
4 – ИБ ФИЦ Коми НЦ УрО РАН (Сыктывкар)
5 – ИОЭБ СО РАН (Улан-Удэ)

CC BY

План

Презентация. Использование данных GBIF - тематика и примеры исследований. 10 минут.

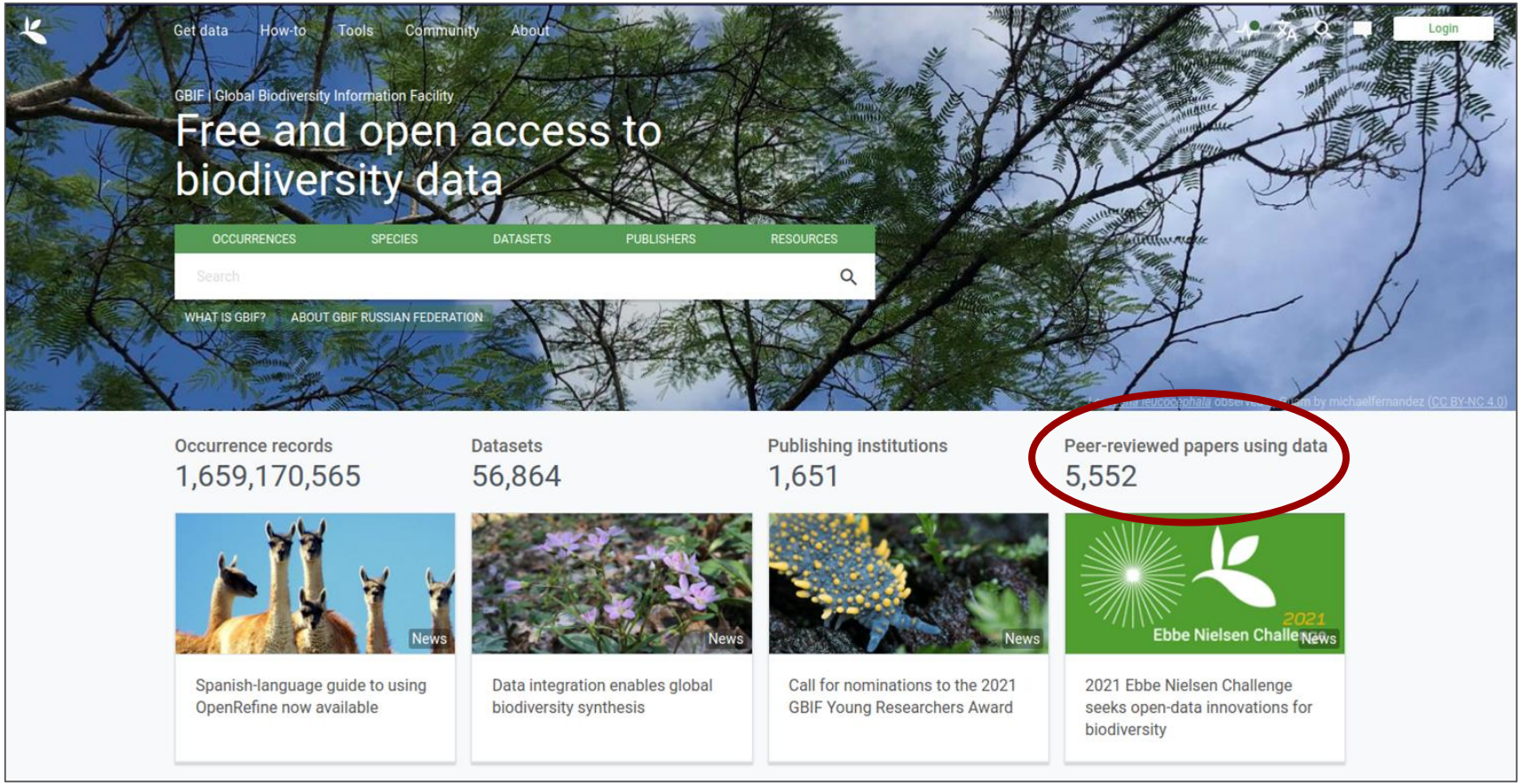
Упражнение. Выгрузка данных из GBIF о находках *Carabus nemoralis* O.F.Müller, 1764. 30 минут.

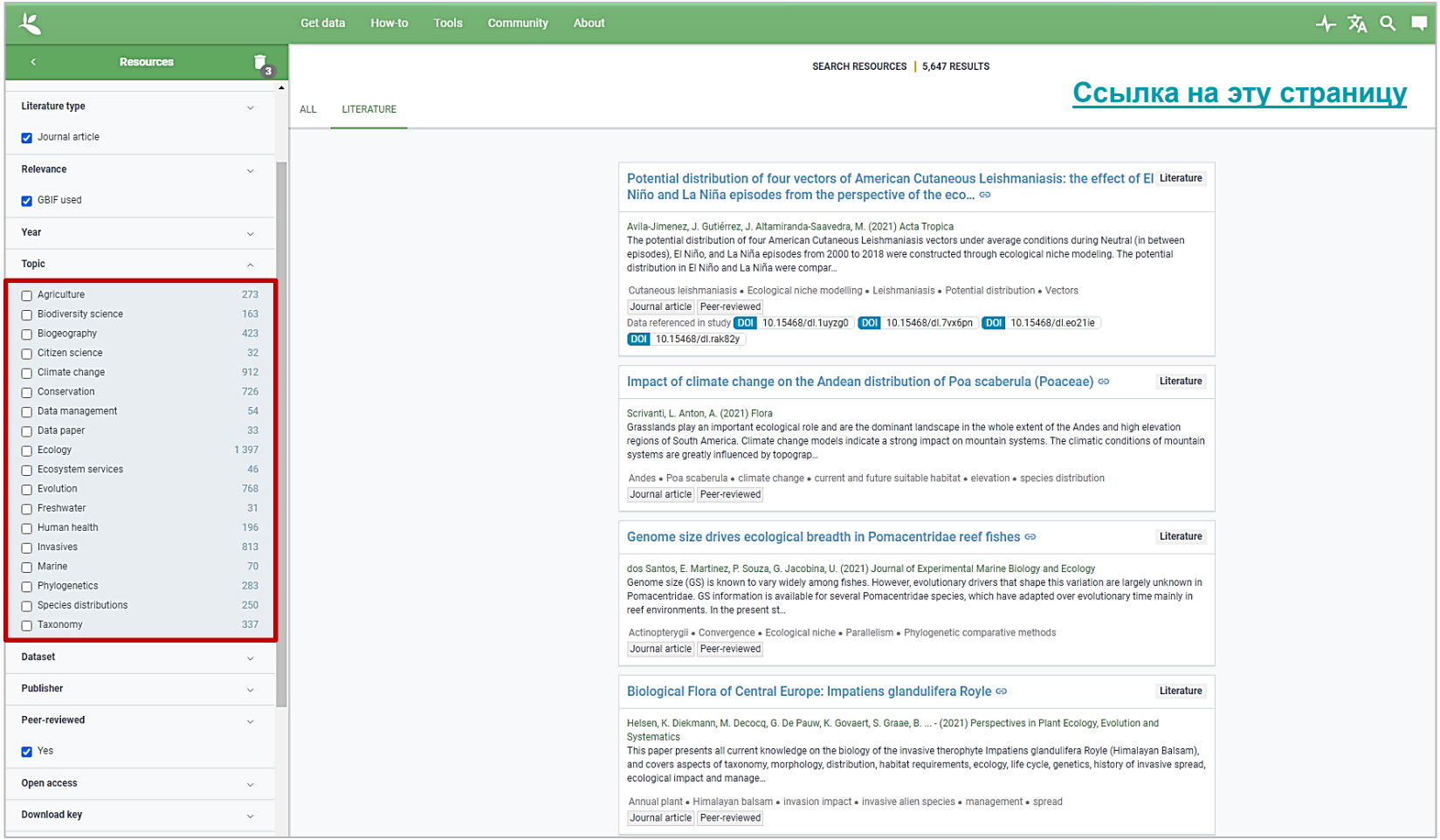
Учебные материалы (доступны на сайте gbif.ru)

- ➔ Как создать личную учетную запись в GBIF
- ➔ Как получить данные, опубликованные в GBIF

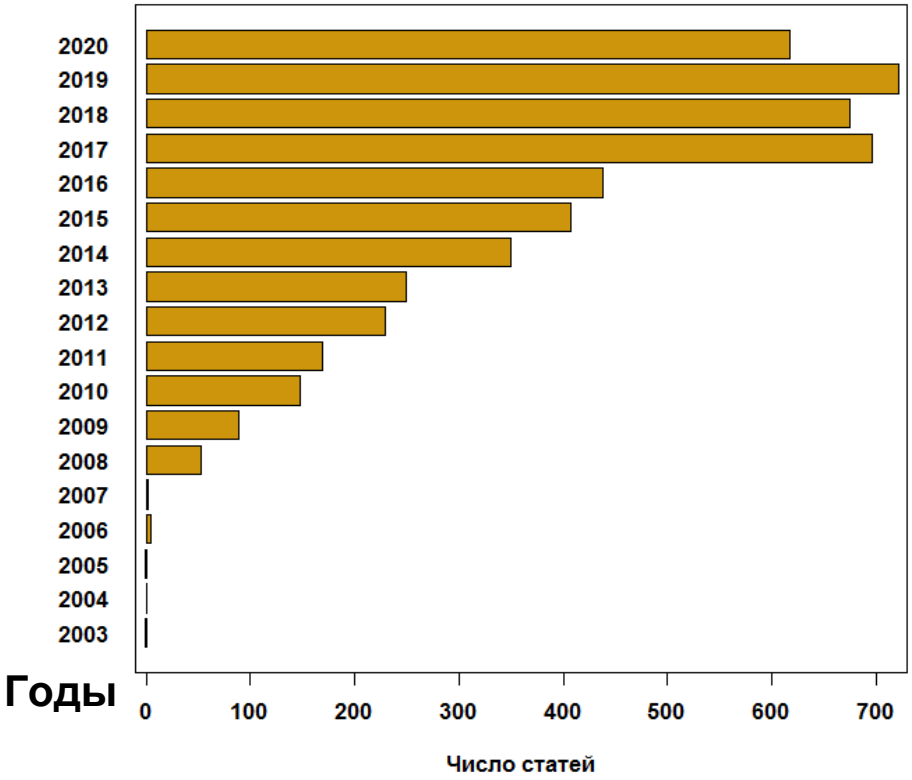
Презентация. Оценка качества данных. Правила цитирования данных GBIF в научных статьях. 20 минут.

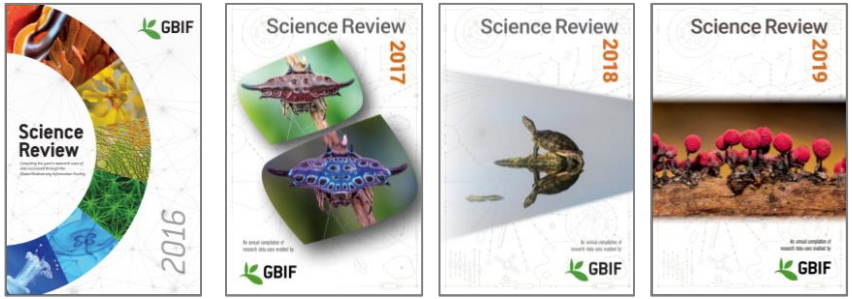
Использование данных GBIF - тематика и примеры исследований


Get data
How-to
Tools
Community
About
Login
GBIF | Global Biodiversity Information Facility
Free and open access to biodiversity data
OCCURRENCES
SPECIES
DATASETS
PUBLISHERS
RESOURCES
Search
WHAT IS GBIF?
ABOUT GBIF RUSSIAN FEDERATION
Occurrence records
1,659,170,565
Datasets
56,864
Publishing institutions
1,651
Peer-reviewed papers using data
5,552
News
Spanish-language guide to using OpenRefine now available
Data integration enables global biodiversity synthesis
Call for nominations to the 2021 GBIF Young Researchers Award
2021
Ebbe Nielsen Challenge
2021 Ebbe Nielsen Challenge seeks open-data innovations for biodiversity


Get data
How-to
Tools
Community
About
Resources
3
Literature type
Journal article
Relevance
GBIF used
Year
Topic
Agriculture 273
Biodiversity science 163
Biogeography 423
Citizen science 32
Climate change 912
Conservation 726
Data management 54
Data paper 33
Ecology 1 397
Ecosystem services 46
Evolution 768
Freshwater 31
Human health 196
Invasives 813
Marine 70
Phylogenetics 283
Species distributions 250
Taxonomy 337
Dataset
Publisher
Peer-reviewed
Yes
Open access
Download key
SEARCH RESOURCES | 5,647 RESULTS
ALL
LITERATURE
Ссылка на эту страницу
Potential distribution of four vectors of American Cutaneous Leishmaniasis: the effect of El Niño and La Niña episodes from the perspective of the eco…
Literature
Avila-Jimenez, J. Gutiérrez, J. Altamiranda-Saavedra, M. (2021) Acta Tropica
The potential distribution of four American Cutaneous Leishmaniasis vectors under average conditions during Neutral (in between episodes), El Niño, and La Niña episodes from 2000 to 2018 were constructed through ecological niche modeling. The potential distribution in El Niño and La Niña were compar…
Cutaneous leishmaniasis • Ecological niche modelling • Leishmaniasis • Potential distribution • Vectors
Journal article
Peer-reviewed
Data referenced in study DOI 10.15468/dl.1uyzg0 DOI 10.15468/dl.7vx6pn DOI 10.15468/dl.eo21ie DOI 10.15468/dl.rak82y
Impact of climate change on the Andean distribution of Poa scaberula (Poaceae)
Literature
Scrivanti, L. Anton, A. (2021) Flora
Grasslands play an important ecological role and are the dominant landscape in the whole extent of the Andes and high elevation regions of South America. Climate change models indicate a strong impact on mountain systems. The climatic conditions of mountain systems are greatly influenced by topograp…
Andes • Poa scaberula • climate change • current and future suitable habitat • elevation • species distribution
Journal article
Peer-reviewed
Genome size drives ecological breadth in Pomacentridae reef fishes
Literature
dos Santos, E. Martinez, P. Souza, G. Jacobina, U. (2021) Journal of Experimental Marine Biology and Ecology
Genome size (GS) is known to vary widely among fishes. However, evolutionary drivers that shape this variation are largely unknown in Pomacentridae. GS information is available for several Pomacentridae species, which have adapted over evolutionary time mainly in reef environments. In the present st…
Actinopterygii • Convergence • Ecological niche • Parallelism • Phylogenetic comparative methods
Journal article
Peer-reviewed
Biological Flora of Central Europe: Impatiens glandulifera Royle
Literature
Helsen, K. Diekmann, M. Decocq, G. De Pauw, K. Govaert, S. Graae, B. ... - (2021) Perspectives in Plant Ecology, Evolution and Systematics
This paper presents all current knowledge on the biology of the invasive therophyte Impatiens glandulifera Royle (Himalayan Balsam), and covers aspects of taxonomy, morphology, distribution, habitat requirements, ecology, life cycle, genetics, history of invasive spread, ecological impact and manage…
Annual plant • Himalayan balsam • invasion impact • invasive alien species • management • spread
Journal article
Peer-reviewed

GBIF science review: самые интересные исследования, выполненные на основе данных GBIF





Обзоры можно посмотреть здесь

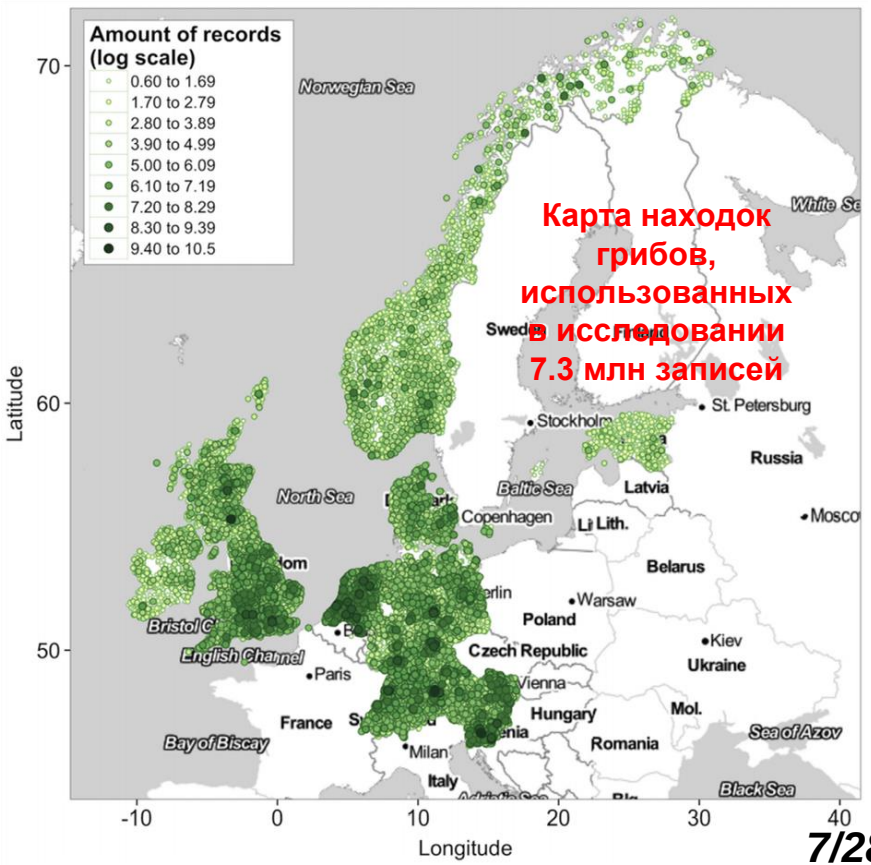
Big data integration: Pan-European fungal species observations assembly for addressing contemporary questions in ecology and global change biology

- 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

Technical Focus

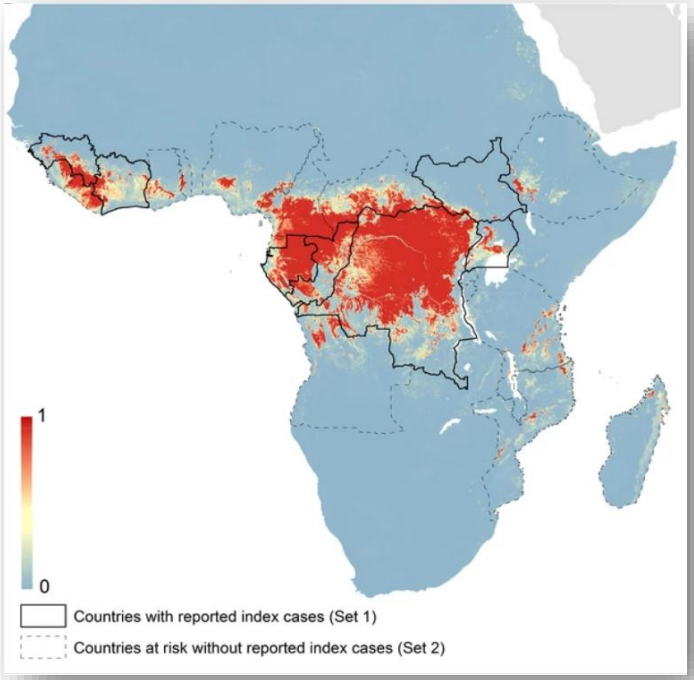
Big data integration: Pan-European fungal species observations' assembly for addressing contemporary questions in ecology and global change biology

Andrew et al.
Fungal Biology Reviews, 2017
https://doi.org/10.1016/j.fbr.2017.01.001



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

Pigott et al., 2014
https://doi.org/10.7554/eLife.04395.001

Упражнение

Выгрузка данных из GBIF
о находках *Carabus nemoralis* O.F.Müller, 1764

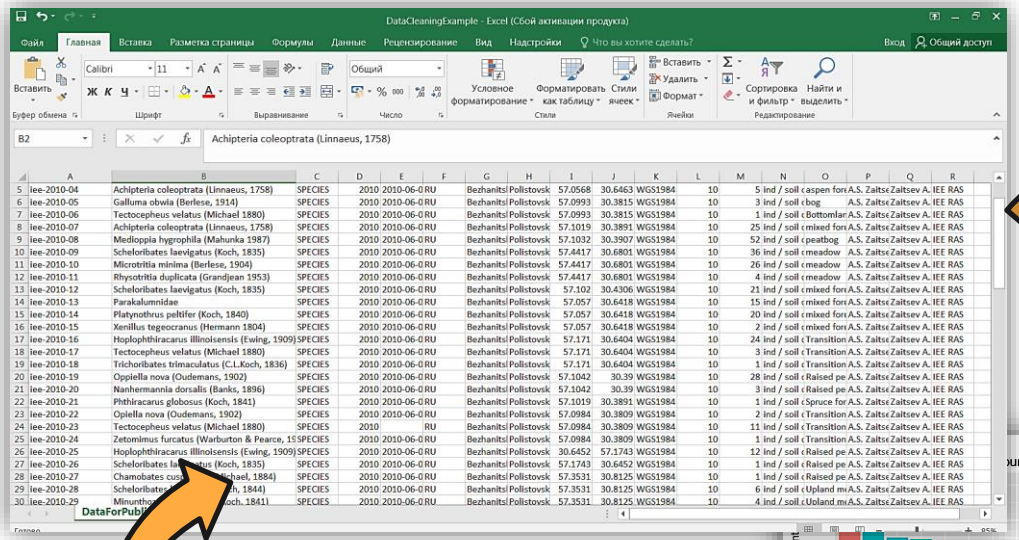
- Создайте поисковые запросы о находках *Carabus nemoralis* O.F.Müller, 1764 в мире и в России, следуя инструкциям. При необходимости создайте учетную запись пользователя.

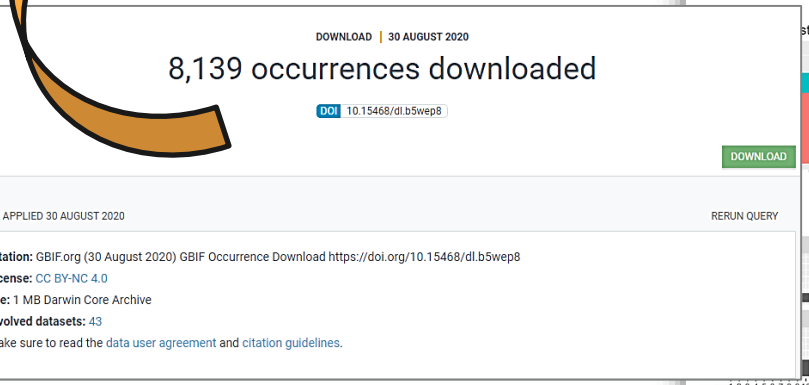
- Обратите внимание на пространственное распределение находок. Какие выводы можно сделать из имеющихся данных?

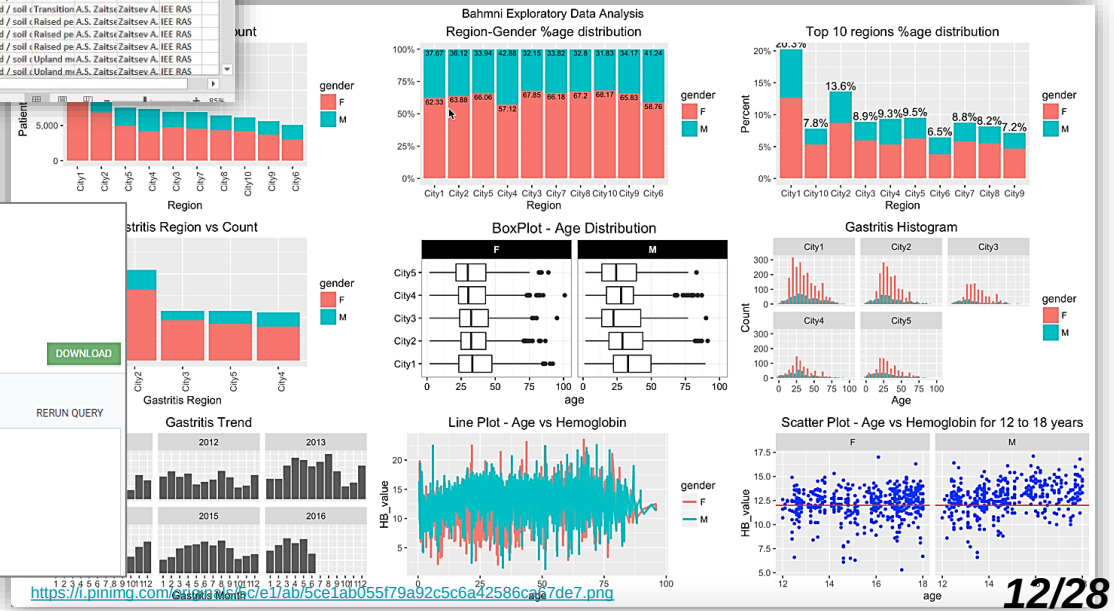
- Выгрузите данные о находках *Carabus nemoralis* O.F.Müller, 1764 в мире и в России в формате SIMPLE. Результаты представьте в таблице:

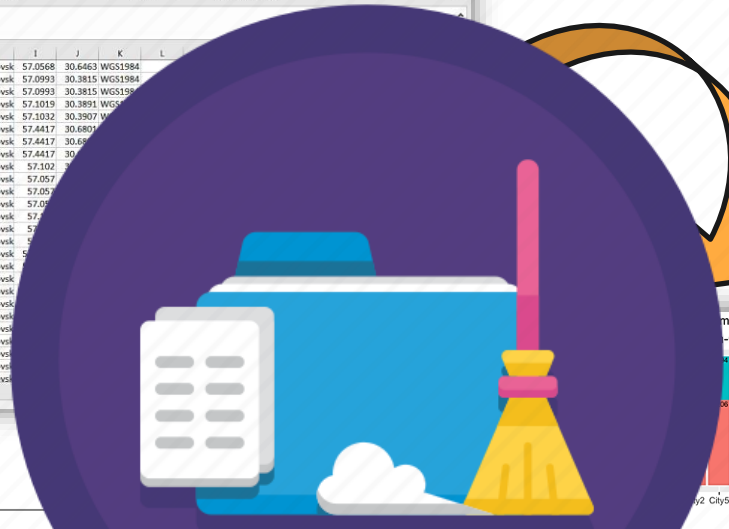
	Находки в мире	**Находки в России**
DOI поискового запроса		
Число записей		
Из каких наборов данных происходят сведения? Перечислите 3 самых больших		

Оценка качества данных


DataCleaningExample - Excel (Сбой активации продукта)
Achipteria coleoptrata (Linnaeus, 1758)


DOWNLOAD | 30 AUGUST 2020
8,139 occurrences downloaded
DOI 10.15468/dl.b5wep8
DOWNLOAD
FILTER APPLIED 30 AUGUST 2020
RERUN QUERY
Citation: GBIF.org (30 August 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.b5wep8
License: CC BY-NC 4.0
File: 1 MB Darwin Core Archive
Involved datasets: 43
Make sure to read the data user agreement and citation guidelines.


Bahmni Exploratory Data Analysis
Region-Gender %age distribution
Top 10 regions %age distribution
Gastritis Region vs Count
BoxPlot - Age Distribution
Gastritis Histogram
Gastritis Trend
Line Plot - Age vs Hemoglobin
Scatter Plot - Age vs Hemoglobin for 12 to 18 years
https://i.pinimg.com/originals/5c/e1/ab/5ce1ab055f79a92c5c6a42586ca97de7.png



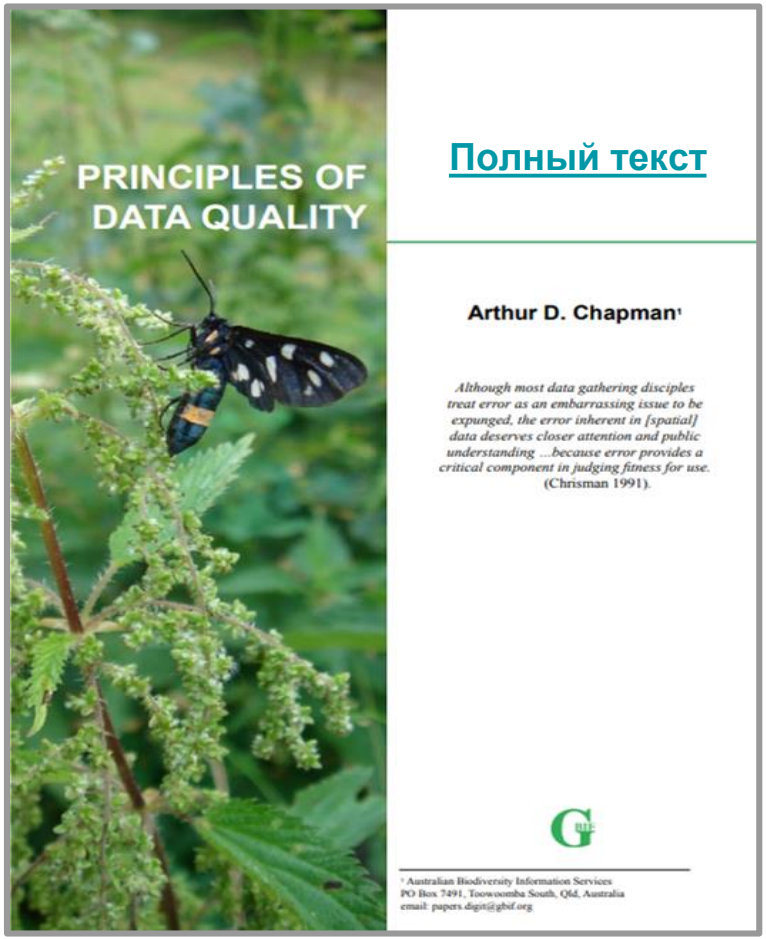
https://i.pinimg.com/originals/5c/e1/ab/5ce1ab055f79a92c5c6a42586ca27de7.png

Данные всегда содержат ошибки

Качество данных – это относительная концепция, которая зависит от способа использования этих данных

Качество данных: можно ли применить полученные через GBIF данные в вашем исследовании?

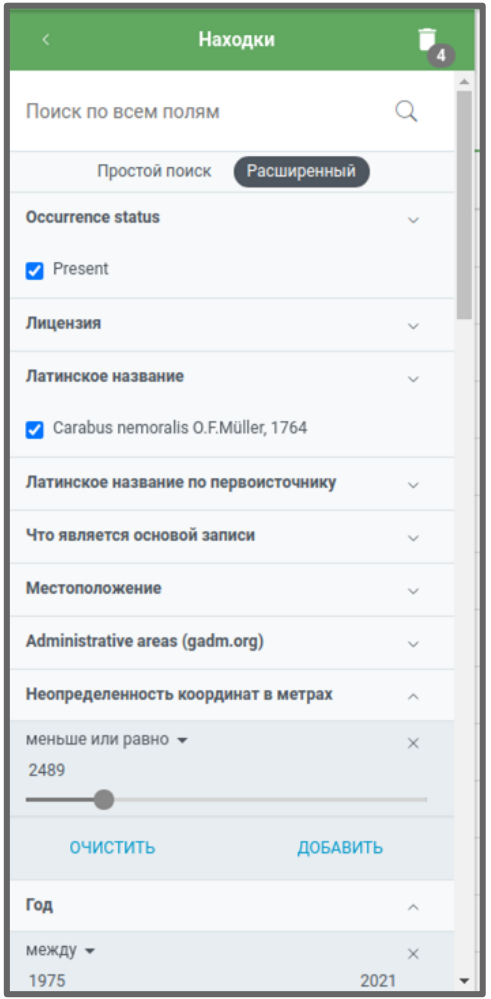
Arthur Chapman
Фото https://github.com/ArthurChapman



Как получить качественные данные?

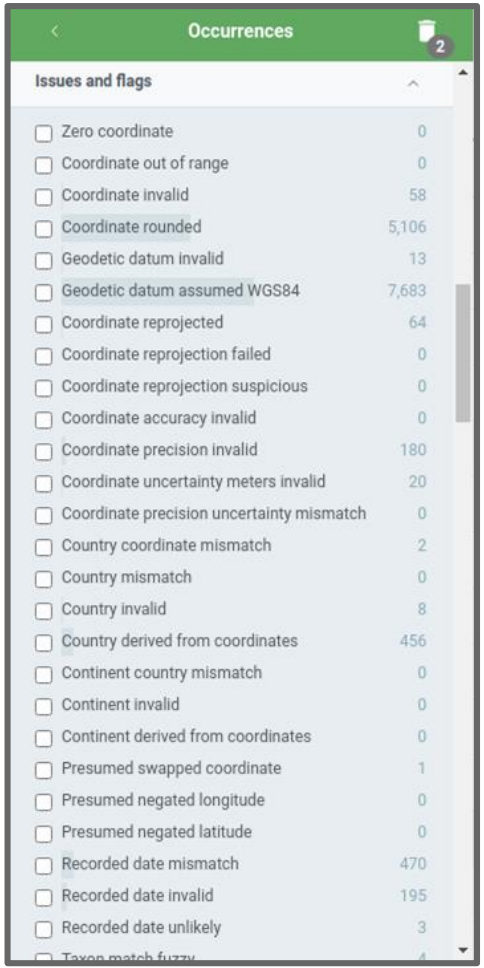
Задайте детальные параметры поиска!

Вы можете отфильтровать ненужные или подозрительные записи до выгрузки данных на свой компьютер.



Как получить качественные данные?

Обратите внимание на автоматически обнаруженные несоответствия



Инструменты для поиска и исправления технических ошибок и ошибок в данных

Текстовые редакторы

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

- R (командная строка)
- RStudio (графический пользовательский интерфейс)

Выберите то, что удобно вам!

Правила цитирования данных GBIF в научных статьях

Почему важно правильно цитировать данные, полученные через GBIF?

Данные сами по себе являются научным продуктом

Для этого данные должны быть:

Легко обнаружимыми через Интернет

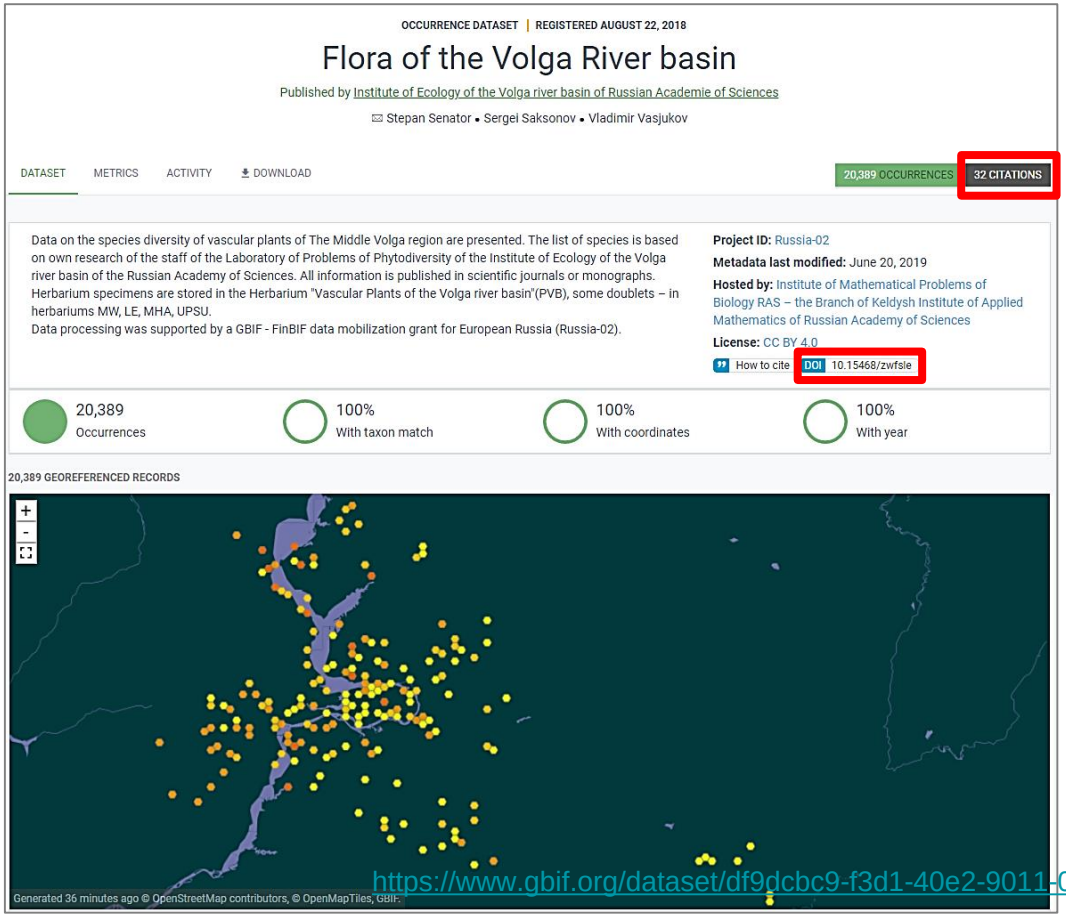
Доступными для исследователей

Совместимыми друг с другом

Многократно используемыми
! Однозначное указание авторства
! Правила использования данных (лицензия)

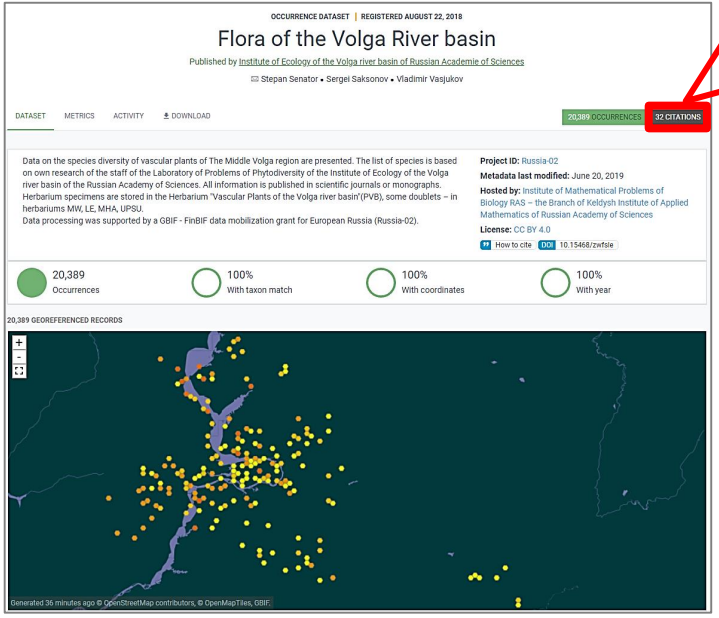
https://doi.org/10.1038/sdata.2016.18

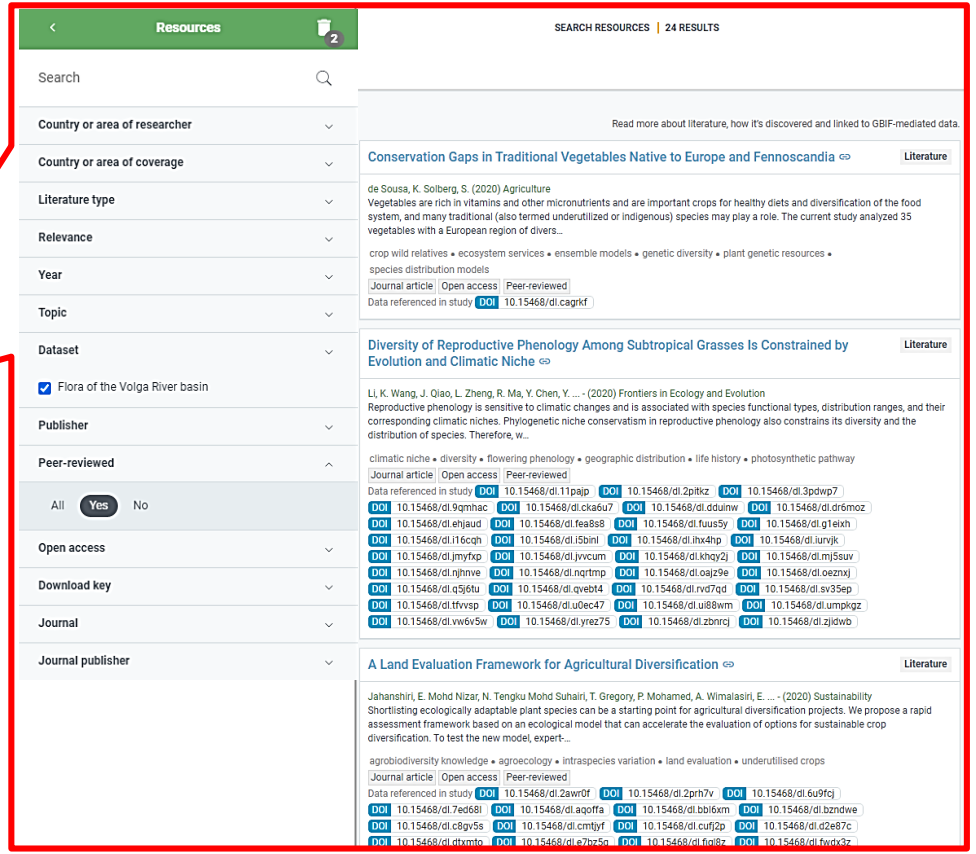
Цитирование данных, доступных через GBIF, по DOI



https://www.gbif.org/dataset/df9dcbc9-f3d1-40e2-9011-02babb0d98f6

Цитирование данных, доступных через GBIF, по DOI





Get data How-to Tools Community About

Citation guidelines

These guidelines provide the most common examples of citation by GBIF users.

Sunfly (*Helophilus pendulus*) on Chicory (*Cichorium intybus*) by Donald Hobern. Photo licensed under CC BY 4.0.

https://www.gbif.org/citation-guidelines

Правила цитирования данных, доступных через GBIF

Доступ к данным через сеть GBIF предоставляется бесплатно для всех, но не без обязательств. В соответствии с условиями пользовательского соглашения, пользователи, которые загружают отдельные наборы данных или результаты поиска и используют их в исследованиях, соглашаются цитировать их, используя DOI.

Пример

Senator S, Saksonov S, Vasjukov V (2018). Flora of the Volga River basin. Version 1.2. Institute of Ecology of the Volga river basin of Russian Academy of Sciences. Occurrence dataset **DOI: 10.15468/zwfsle** accessed via GBIF.org on 2020-09-16.

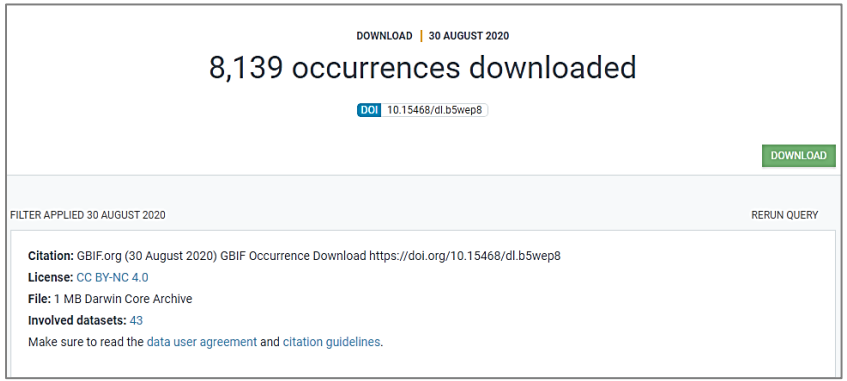
Мы использовали данные из gbif.org = **Мы использовали методику, опубликованную в журнале «Экология»**

Почему важно правильно цитировать данные, полученные через GBIF?

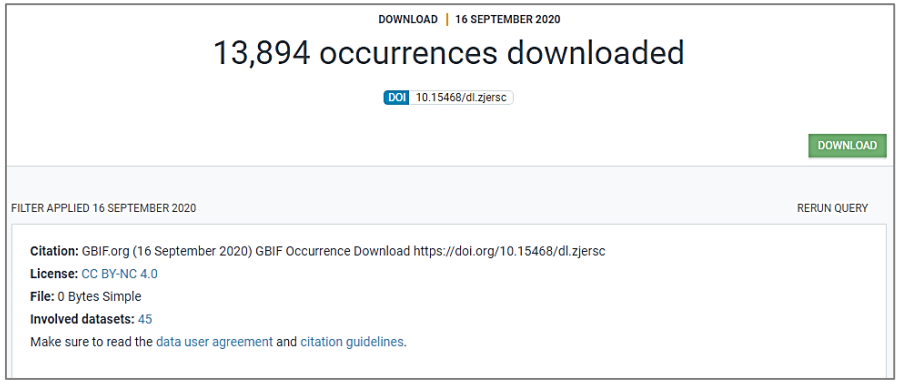
Данные в GBIF постоянно обновляются

Пример. Находки жужелиц (Carabidae) для территории России

30 августа 2020 г. **8 139** записей



16 сентября 2020 г. **13 894** записи



GBIF.org (30 August 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.b5wep8

GBIF.org (16 September 2020) GBIF Occurrence Download https://doi.org/10.15468/dl.zjersc

Заключение

- Данные, доступные через GBIF, могут быть важным источником информации в вашем исследовании.
- Для получения качественных данных важно составить подробный поисковый запрос.
- Любые данные перед анализом нуждаются в критической оценке.
- Загружая данные через GBIF, вы обязуетесь при их использовании соблюдать условия Пользовательского соглашения.
- На данные необходимо ссылаться в списке литературы, также как и на другие источники (статьи, монографии, тезисы и др.)
- Правильное цитирование данных, полученных через GBIF, повышает их статус как надежного источника научной информации.

Домашнее задание

Создайте поисковый запрос о находках *Carabus nemoralis* O.F.Müller, 1764 по следующим критериям:

- Пространственный охват – весь мир.
- Только образцы из коллекций (меню поиска Basis of record).
- Записи, для которых страна сбора установлена по координатам (меню поиска Issues and flags), необходимо исключить.

Выгрузите данные в формате Darwin Core archive.

Ответьте на вопросы:

- Из каких наборов данных происходят полученные сведения?
- По какой лицензии распространяются выгруженные данные?
- Как ссылаться на полученные данные в научных статьях?

Выполненное задание вы можете прислать на адрес gbif.ru@yandex.ru до 15 апреля

Естественнонаучные коллекции Югры: сбор, фиксация, хранение, введение в научный оборот
Ханты-Мансийск, 30 марта — 1 апреля 2021

Использование данных GBIF в научных исследованиях: введение

gbif.ru@yandex.ru

Спасибо за внимание!

CC BY