



ЛЕКЦИЯ 8

Использование массивов объединенных данных в научных исследованиях. Лицензии и авторские права. Качество данных.

Слайды CC BY:

Paula Zermoglio, John Wieczorek
Наталья Иванова,
Максим Шашков



План лекции

Лицензии на данные и авторские права в GBIF

Что такое “качество данных”

Основные типы ошибок в данных и способы их исправления

Документирование

Контроль качества данных порталом GBIF

Таксономический справочник GBIF Backbone

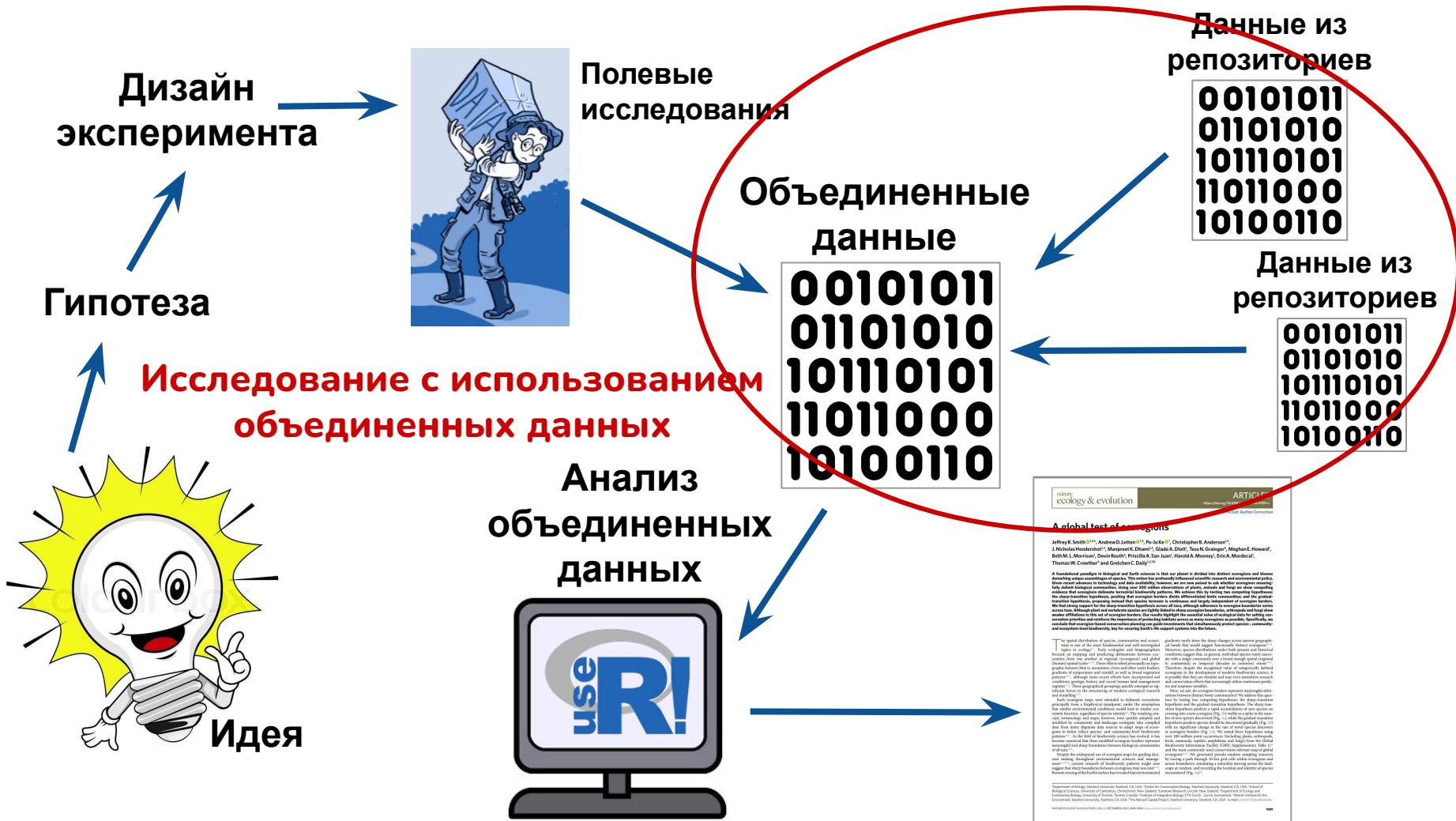


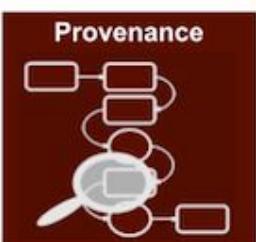
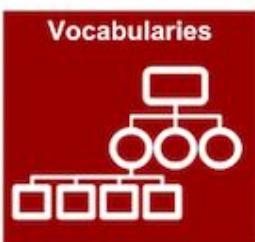
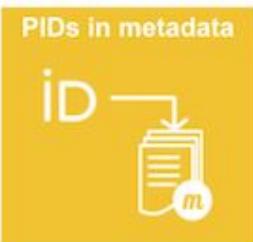
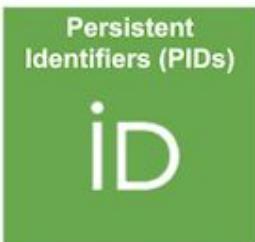
Arthur D. Chapman¹

Although most data gathering disciplines treat error as an embarrassing issue to be expunged, the error inherent in [spatial] data deserves closer attention and public understanding ...because error provides a critical component in judging fitness for use.
(Chrisman 1991).

GBIF

¹ Australian Biodiversity Information Services
PO Box 7491, Toowoomba South, Qld, Australia
email: papers.digit@gbif.org







Лицензии Creative Commons

Лицензия CC-0 позволяет другим распространять, редактировать, поправлять и брать за основу ваше произведение, даже коммерчески.

Лицензия CC-BY позволяет другим распространять, редактировать, поправлять и брать за основу ваше произведение, даже коммерчески, с обязательным указанием авторства.

Лицензия CC-BY-NC позволяет другим изменять, поправлять и брать за основу ваше произведение некоммерческим образом. Обязательно указывать авторство.

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)
by 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 m² (for a t... [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

5,693 Occurrences

100% With taxon match

100% With coordinates

100% With year

5,693 GEOREFERENCED 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



DOWNLOAD | 16 APRIL 2022

2,543 occurrences included in download

DOI 10.15468/dl.qzgpn2

DOWNLOAD

PLEASE USE THIS CITATION IN PUBLICATIONS

GBIF.org (16 April 2022) GBIF Occurrence Download <https://doi.org/10.15468/dl.qzgpn2>

[Copy](#) [BibTex](#) [RIS](#)

FILTER APPLIED 16 APRIL 2022

RERUN QUERY

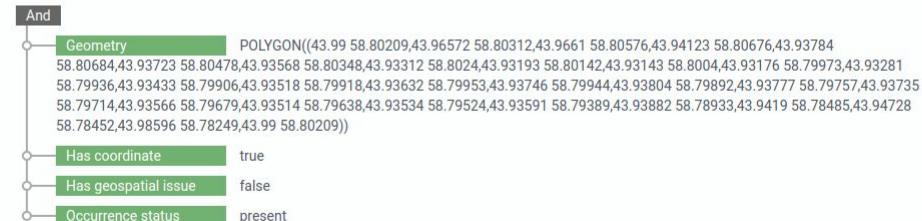
Licence: CC BY-NC 4.0

File: 282 KB Darwin Core Archive

Involved datasets: 6

Make sure to read the [data user agreement](#) and [citation guidelines](#).

API

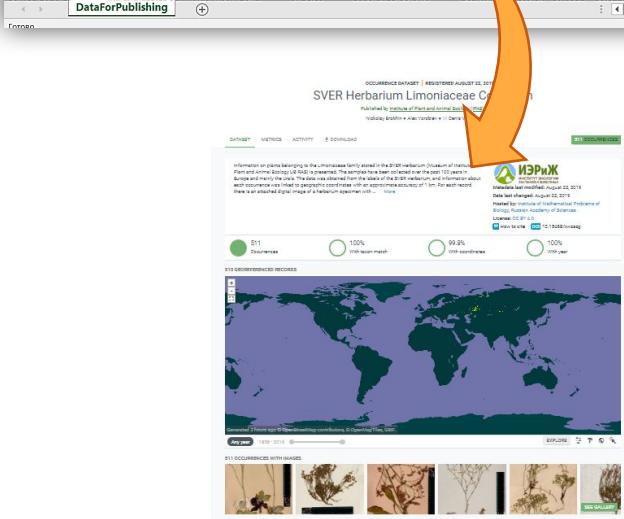


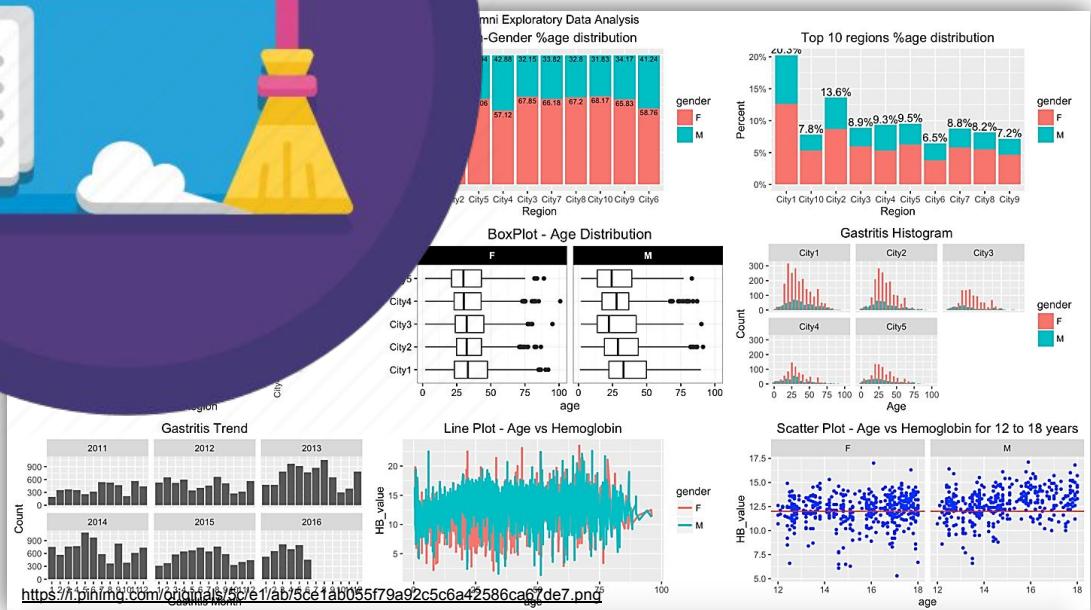
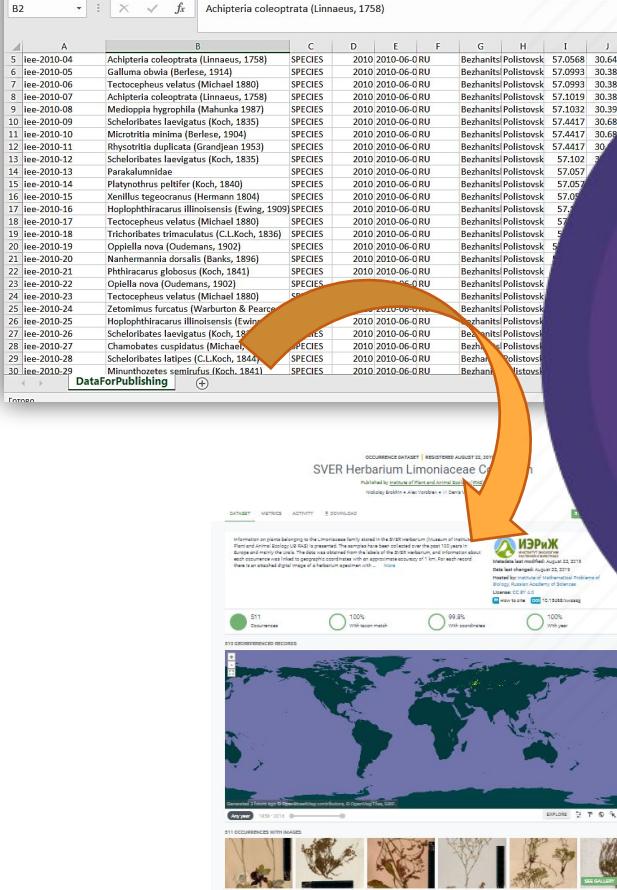
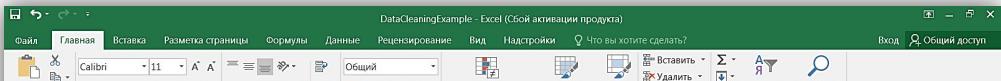


Лицензия для данных поискового запроса - самая строгая из составляющих его наборов данных

B2 A Achipteria coleoptrata (Linnaeus, 1758)

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O	P	Q	R
5	lee-2010-04	Achipteria coleoptrata (Linnaeus, 1758)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0568	30.6463	WG51984	10	5 Ind / soil : aspen for A.S. Zaitsev A.IEE RAS						
6	lee-2010-05	Gallium obvia (Linné, 1753)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0994	30.3861	WG51984	10	3 Ind / soil : bog						
7	lee-2010-06	Tetracepheus velutinus (Michael 1860)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0994	30.3861	WG51984	10	1 Ind / soil : Bog						
8	lee-2010-07	Achipteria coleoptrata (Linnaeus, 1758)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1039	30.8891	WG51984	10	25 Ind / soil : aspen for A.S. Zaitsev A.IEE RAS						
9	lee-2010-08	Medicago lupulina (Linné, 1753)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1039	30.3907	WG51984	10	52 Ind / soil : flowering						
10	lee-2010-09	Schelorhiza laevigata (Koch, 1835)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.4417	30.6801	WG51984	10	36 Ind / soil : meadow						
11	lee-2010-10	Microtritia minima (Bergse, 1904)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.4417	30.6801	WG51984	10	26 Ind / soil : meadow						
12	lee-2010-11	Rhynchospora duplex (Grandjean 1953)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.4417	30.6801	WG51984	10	4 Ind / soil : meadow						
13	lee-2010-12	Schelorhiza laevigata (Koch, 1835)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.102	30.4306	WG51984	10	21 Ind / soil : mixed for A.S. Zaitsev A.IEE RAS						
14	lee-2010-13	Parakalmiidae	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.057	30.6418	WG51984	10	15 Ind / soil : mixed for A.S. Zaitsev A.IEE RAS						
15	lee-2010-14	Platynothrus peltifer (Koch, 1840)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.057	30.6418	WG51984	10	20 Ind / soil : mixed for A.S. Zaitsev A.IEE RAS						
16	lee-2010-15	Xerinx tegeocerasus (Hermann 1804)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.057	30.6418	WG51984	10	2 Ind / soil : mixed for A.S. Zaitsev A.IEE RAS						
17	lee-2010-16	Hedysarum occidentale (Ewing, 1909)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.171	30.6404	WG51984	10	24 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
18	lee-2010-17	Tetracepheus velutinus (Michael 1860)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.171	30.6404	WG51984	10	3 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
19	lee-2010-18	Trichorhiza tricuspidata (C.L.Koch, 1836)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.171	30.6404	WG51984	10	1 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
20	lee-2010-19	Oppelia nova (Oudemans, 1902)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1042	30.39	WG51984	10	28 Ind / soil : Raised pe A.S. Zaitsev A.IEE RAS						
21	lee-2010-20	Nanthermannia dorsalis (Banks, 1896)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1042	30.39	WG51984	10	3 Ind / soil : Raised pe A.S. Zaitsev A.IEE RAS						
22	lee-2010-21	Phthiranthes globosa (Koch, 1841)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1019	30.3891	WG51984	10	1 Ind / soil : Spruce for A.S. Zaitsev A.IEE RAS						
23	lee-2010-22	Opella nova (Oudemans, 1902)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0984	30.3809	WG51984	10	2 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
24	lee-2010-23	Tectocarpus velutinus (Michael 1860)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0984	30.3809	WG51984	10	11 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
25	lee-2010-24	Zetomimus furcatus (Warburton & Pearce 1895)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.0984	30.3809	WG51984	10	1 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
26	lee-2010-25	Hoplophyllum illinoense (Ewing, 1909)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	56.652	30.1743	WG51984	10	12 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
27	lee-2010-26	Schelorhiza laevigata (Koch, 1835)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.1743	30.6452	WG51984	10	1 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
28	lee-2010-27	Chloranthus spiculatus (Michael 1860)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.3531	30.8125	WG51984	10	6 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
29	lee-2010-28	Schelorhiza latipes (C.L.Koch, 1844)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.3531	30.8125	WG51984	10	4 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						
30	lee-2010-29	Minuartia venutiflora (Koch, 1841)	SPECIES	2010	2010-06-0RU	Bezhanitsi	Polisskoy	57.3531	30.8125	WG51984	10	1 Ind / soil : Transition A.S. Zaitsev A.IEE RAS						

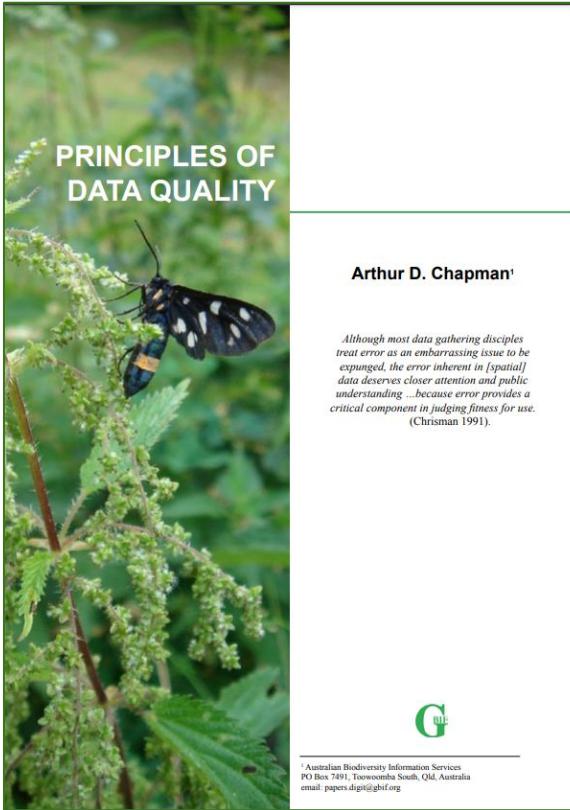




Данные всегда содержат ошибки

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





Основные подходы к оценке качества данных



Chapman AD (2005) Principles of Data Quality. Global Biodiversity Information Facility.
<https://doi.org/10.15468/doc.jrgg-a190>

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)

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



Выберите то, что удобно вам!



аудитор данных



автор

Графическая форма удобная для чтения
неудобна для передачи информации и хранения

	00	01	02	03	04	05	06	07	08	09	0A	0B	0C	0D	0E	0F
00	NUL	STX	SOT	ETX	EOT	ENQ	ACK	BEL	BS	HT	LF	VT	FF	CR	SO	SI
0000	0001	0002	0003	0004	0005	0006	0007	0008	0009	000A	000B	000C	000D	000E	000F	
10	DLE	DC1	DC2	DC3	DC4	NAK	SYN	ETB	CAN	EM	SUB	ESC	FS	GS	RS	US
0010	0011	0012	0013	0014	0015	0016	0017	0018	0019	001A	001B	001C	001D	001E	001F	
20	SP	!	"	#	\$	%	&	*	()	*	+	,	-	.	/
0020	0021	0022	0023	0024	0025	0026	0027	0028	0029	002A	002B	002C	002D	002E	002F	
30	0	1	2	3	4	5	6	7	8	9	:	;	<	=	>	?
0030	0031	0032	0033	0034	0035	0036	0037	0038	0039	003A	003B	003C	003D	003E	003F	
40	Ø	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
0040	0041	0042	0043	0044	0045	0046	0047	0048	0049	004A	004B	004C	004D	004E	004F	
50	P	Q	R	S	T	U	V	W	X	Y	Z	[\]	^	
0050	0051	0052	0053	0054	0055	0056	0057	0058	0059	005A	005B	005C	005D	005E	005F	
60	‘	а	б	с	д	е	ф	г	һ	и	ј	к	л	м	н	օ
0060	0061	0062	0063	0064	0065	0066	0067	0068	0069	006A	006B	006C	006D	006E	006F	
70	р	զ	ր	ս	տ	ս	ւ	ա	չ	յ	շ	{		}	~	DEL
0070	0071	0072	0073	0074	0075	0076	0077	0078	0079	007A	007B	007C	007D	007E	007F	
80	Ђ	Ѓ	Ѓ	Ѓ	Ѓ	Ѓ	Ѓ	Ѓ	Ѓ	Ђ	Ѓ	Ќ	Ћ	Џ	Џ	
0402	0403	201A	0453	201E	2026	2020	2021	20AC	2030	0409	2039	040A	040C	040B	040F	
90	Ђ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ	Ѡ
0452	2018	2019	201C	201D	2022	2013	2014	2015	2016	2122	0459	203A	045A	045C	045B	045F
A0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
00A0	040E	045E	0408	00A4	0490	00A6	00A7	0401	00A9	0404	00AB	00AC	00AD	00AE	0407	
B0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
00B0	00B1	0406	0456	0491	00B5	00B6	00B7	0451	2116	0454	00BB	0458	0405	0455	0457	
C0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
0410	0411	0412	0413	0414	0415	0416	0417	0418	0419	041A	041B	041C	041D	041E	041F	
D0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
0420	0421	0422	0423	0424	0425	0426	0427	0428	0429	042A	042B	042C	042D	042E	042F	
E0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
0430	0431	0432	0433	0434	0435	0436	0437	0438	0439	043A	043B	043C	043D	043E	043F	
F0	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩	₩
0440	0441	0442	0443	0444	0445	0446	0447	0448	0449	044A	044B	044C	044D	044E	044F	

буквы английского (латиница) и
русского (кириллица) алфавитов

E
0415

E
0045

C
0421

C
0043

C - самая "опасная"

Temporagу reservoir 37 km SE Dzhangala
Temporagу reservoir

Старицы р. Урал юго-вост. пос. Есенсай

Старицы р. Урал юго-вост. пос. Есенсай

Старицы р. Урал юго-вост. пос. Есенсай

Пойма реки Урал у поселка Есенсай

Пойма реки Урал у поселка Есенсай

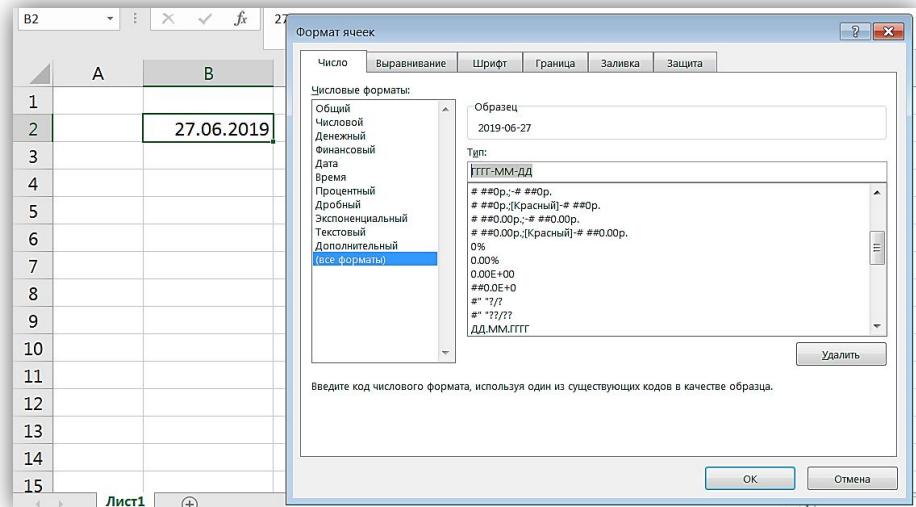
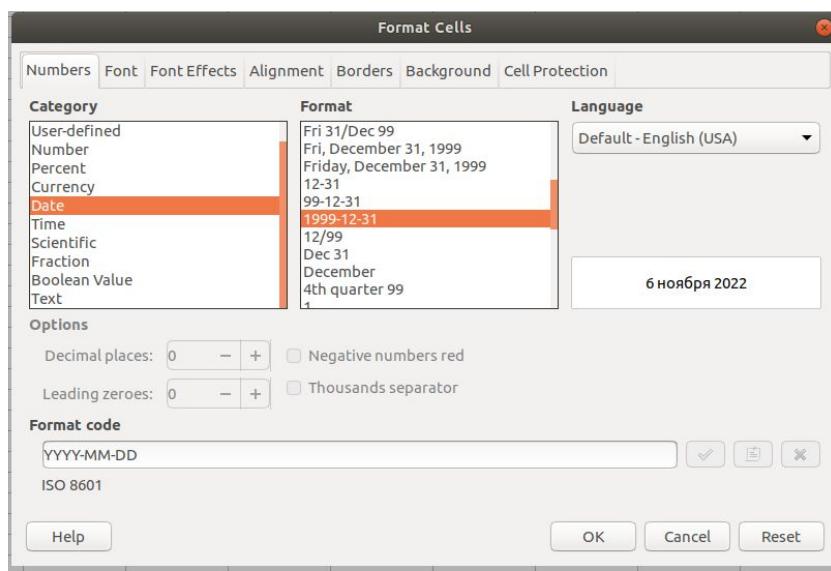
Есенсай

три разных слова!

Есенсай

Есенсай

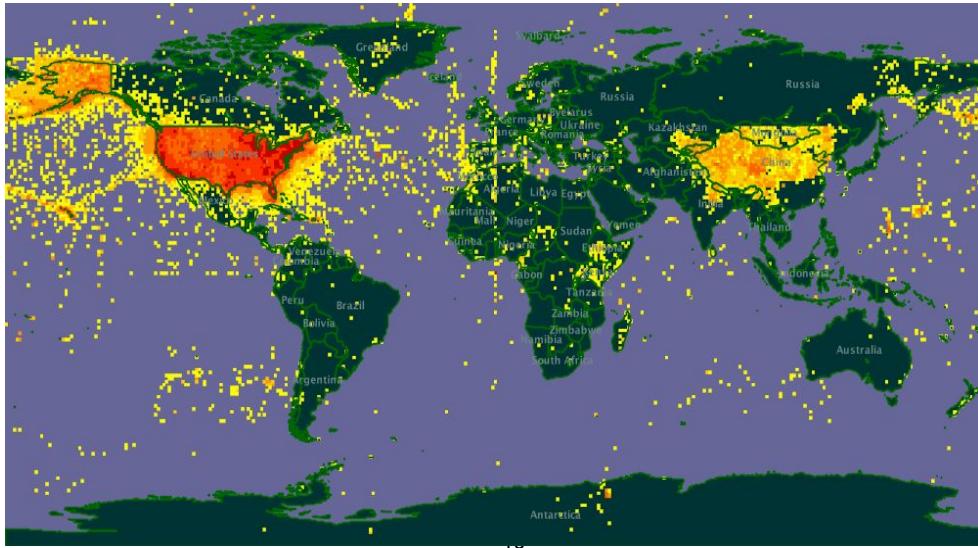
eventDate	verbatimEventDate
2019-08-27	27 авг 2019
2019-08-27	27 VIII 2019
2019-08-27	27.08.2019
2019-08-29/30	29-30 августа 2019



Как правильно задать формат даты ГГГГ-ММ-ДД

Пространственные данные: наиболее распространенные технические ошибки

- 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



Ранняя GBIF карта, иллюстрирующая данные из США, с широко распространенными ошибками:

- Координаты 0,0 (Гринвичский меридиан и Экватор)
- Неправильно указано полушарие (точки с неверной (восточной) долготой попадают в Китай, с неверной (южной) широтой - в Чили).

Пространственные данные: ошибки формата

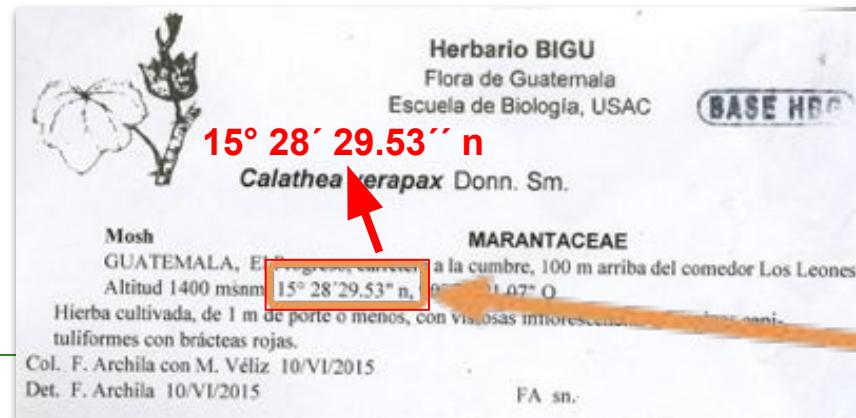
Градусы Минуты Секунды Полушарие →
десятичные градусы

$$\Gamma\Gamma = (\Gamma + M/60 + C/3600) * [\text{Полушарие}]$$

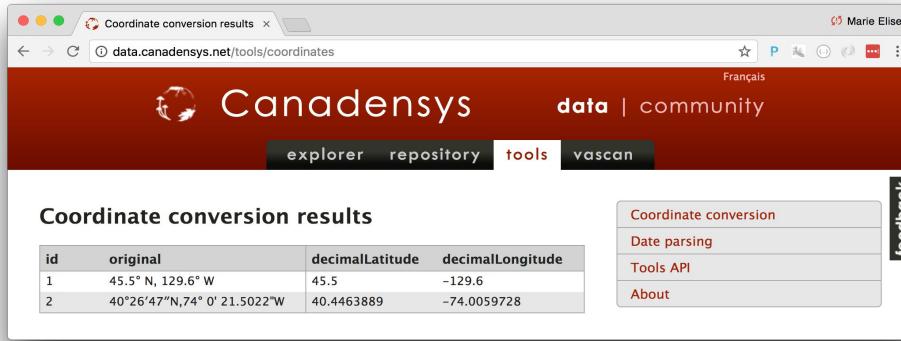
Полушарие: западное = -1; восточное = 1

$$\Gamma\Gamma = (15 + 28/60 + 29.53/3600) * 1$$

$$\Gamma\Gamma = 15.47487$$



Автоматический пересчет координат из ГГ ММ СС в ГГ.ГГГГГ



The screenshot shows a web browser window with the title 'Coordinate conversion results'. The URL is 'data.canadensys.net/tools/coordinates'. The page has a dark red header with the 'Canadensys' logo and 'data | community' text. Below the header, there are tabs for 'explorer', 'repository', 'tools' (which is selected), and 'vascan'. A sidebar on the right contains links for 'Coordinate conversion', 'Date parsing', 'Tools API', and 'About'. The main content area displays a table of coordinate conversion results:

id	original	decimalLatitude	decimalLongitude
1	45° 5' N, 129° 6' W	45.5	-129.6
2	40°26'47"N,74° 0' 21.5022"W	40.4463889	-74.0059728

<http://data.canadensys.net/tools/coordinates?lang=en>



The screenshot shows the 'GIS LAB' website. The header features the 'GIS LAB' logo, social media links for Twitter, Facebook, and Google+, and a language selection dropdown set to 'English'. The main navigation menu includes 'Статьи', 'Документация', 'Геоданные', 'О GIS-Lab', and 'С чего начать?'. Below the menu, there is a breadcrumb trail: 'Главная / Вопросы и ответы'. The main content area has a dark blue sidebar with the text 'Как переводить координаты из одного числового формата в другой'. The main content area displays the following text:

Конвертация значений координат в формате DDMMMSS в формат DD.DDDD

At the bottom of the page, there are buttons for 'Обсудить в форуме' (Discuss in forum) and 'Редактировать в вики' (Edit in wiki).

<http://gis-lab.info/qa/dms2dd.html>

WHAT THE NUMBER OF DIGITS IN YOUR COORDINATES MEANS

LAT/LON PRECISION

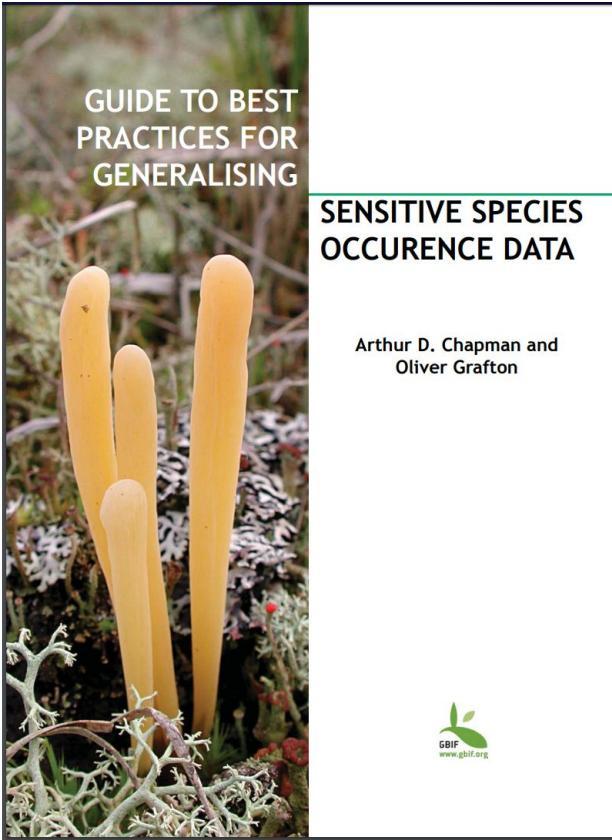
MEANING

28°N, 80°W	YOU'RE PROBABLY DOING SOMETHING SPACE-RELATED
28.5°N, 80.6°W	YOU'RE POINTING OUT A SPECIFIC CITY
28.52°N, 80.68°W	YOU'RE POINTING OUT A NEIGHBORHOOD
28.523°N, 80.683°W	YOU'RE POINTING OUT A SPECIFIC SUBURBAN CUL-DE-SAC
28.5234°N, 80.6830°W	YOU'RE POINTING TO A PARTICULAR CORNER OF A HOUSE
28.52345°N, 80.68309°W	YOU'RE POINTING TO A SPECIFIC PERSON IN A ROOM, BUT SINCE YOU DIDN'T INCLUDE DATUM INFORMATION, WE CAN'T TELL WHO
28.5234571°N, 80.6830941°W	YOU'RE POINTING TO WALDO ON A PAGE
28.523457182°N, 80.683094159°W	"HEY, CHECK OUT THIS SPECIFIC SAND GRAIN!"
28.523457182818284°N, 80.683094159265358°W	EITHER YOU'RE HANDING OUT RAW FLOATING POINT VARIABLES, OR YOU'VE BUILT A DATABASE TO TRACK INDIVIDUAL ATOMS. IN EITHER CASE, PLEASE STOP.

Сколько нужно знаков
после запятой?

<https://xkcd.com/2170/>

Sensitive data - данные, не подлежащие разглашению



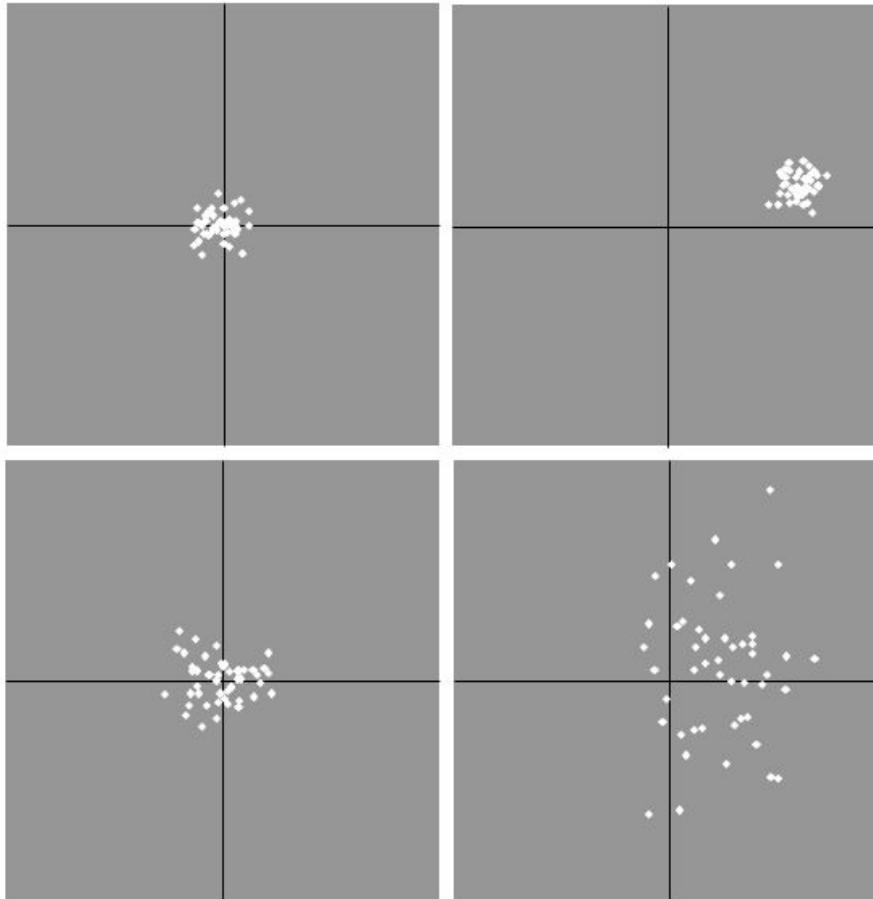
Методика генерализации информации о находках охраняемых видов

Chapman AD & Grafton O (2008) Guide to Best Practices for Generalising Primary Species-Occurrence Data, version 1.0. Copenhagen: Global Biodiversity Information Facility, 27 pp. ISBN: 87-92020-06-2. Available online at <http://www.gbif.org/resource/80512>.

accurate

inaccurate

precise



Оценка погрешностей в определении координат

Accurate - аккуратность определения координат (dwc: coordinateUncertaintyInMeters), погрешность метода

Precise - точность прибора (dwc: coordinatePrecision), погрешность прибора



Georeferencing Best Practices

Arthur D. Chapman, John R. Wieczorek

Version 743e3a1, 2022-02-25 15:06:09 UTC

[Ссылка](#)



Georeferencing Quick Reference Guide

Paula F. Zermoglio, Arthur D. Chapman, John R. Wieczorek, Maria Celeste Luna, David A. Bloom

Version 4ac9d96, 2022-02-25 15:09:32 UTC

[Ссылка](#)

QGIS: открытая ГИС



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



Проверка корректности данных

Массив данных содержит образцы окаменелостей Триасового периода.

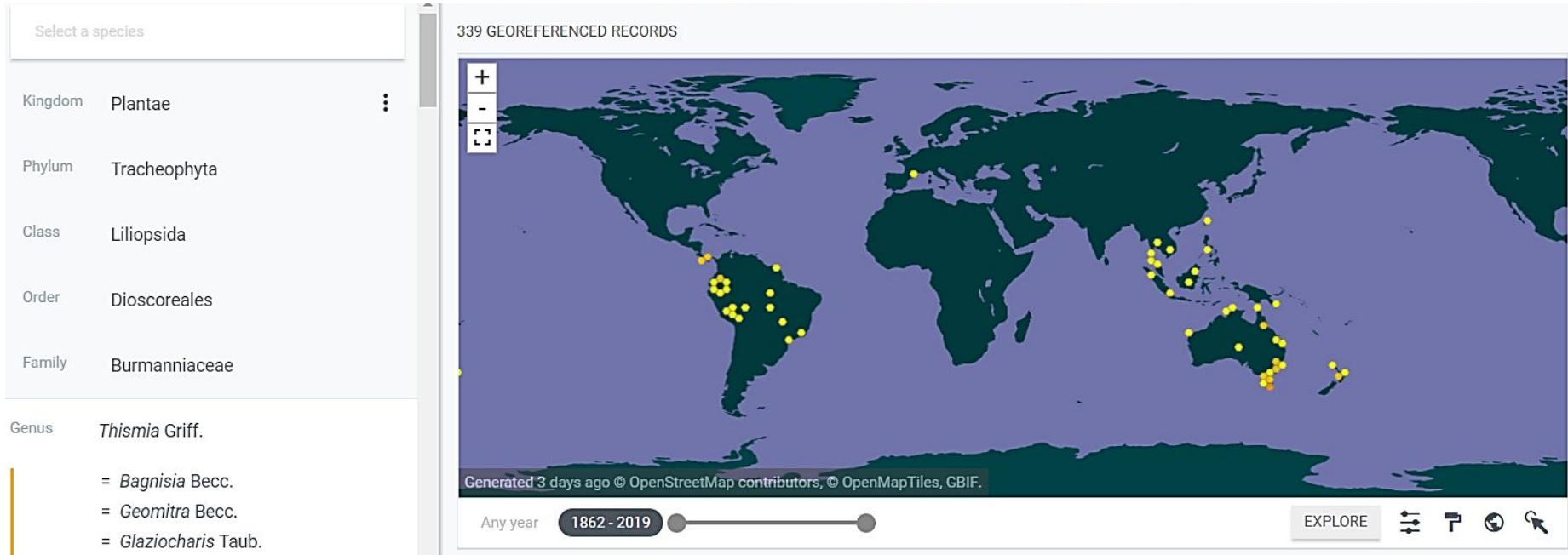
Представлены записи для образца рода *Thismia*.

***Thismia* – это ископаемый вид?**



Проверка корректности данных

Thismia – род современных цветковых растений

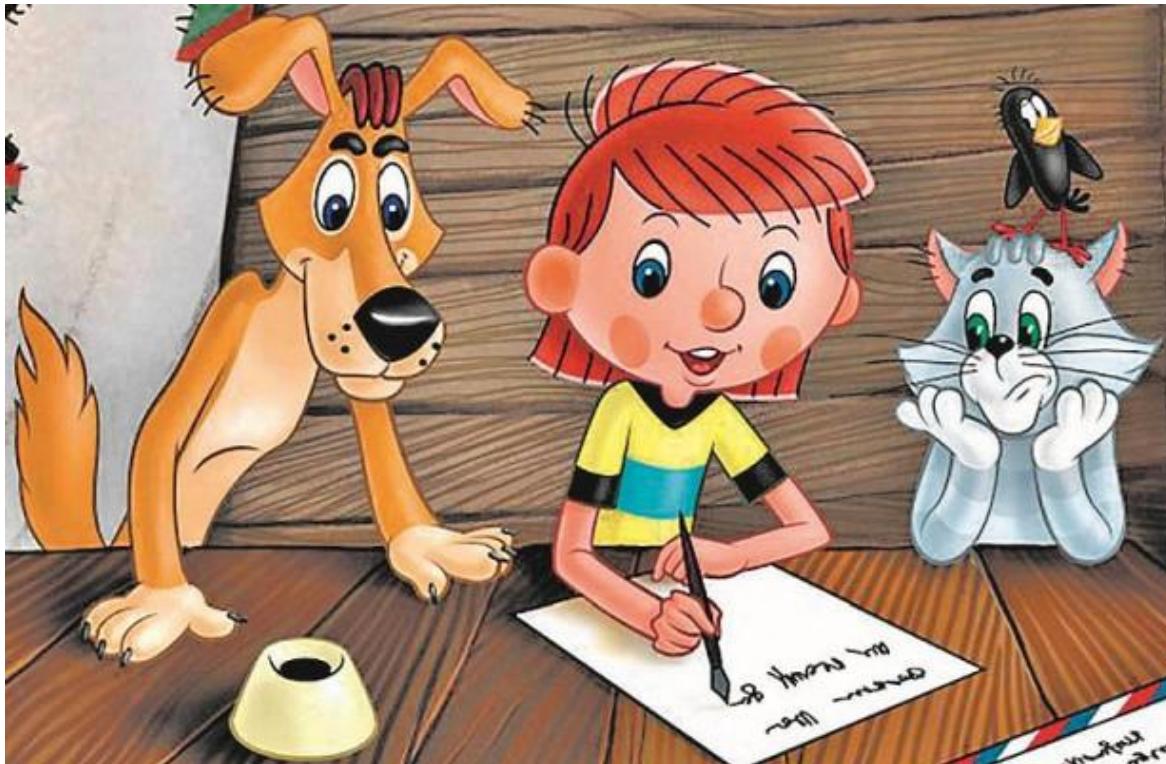


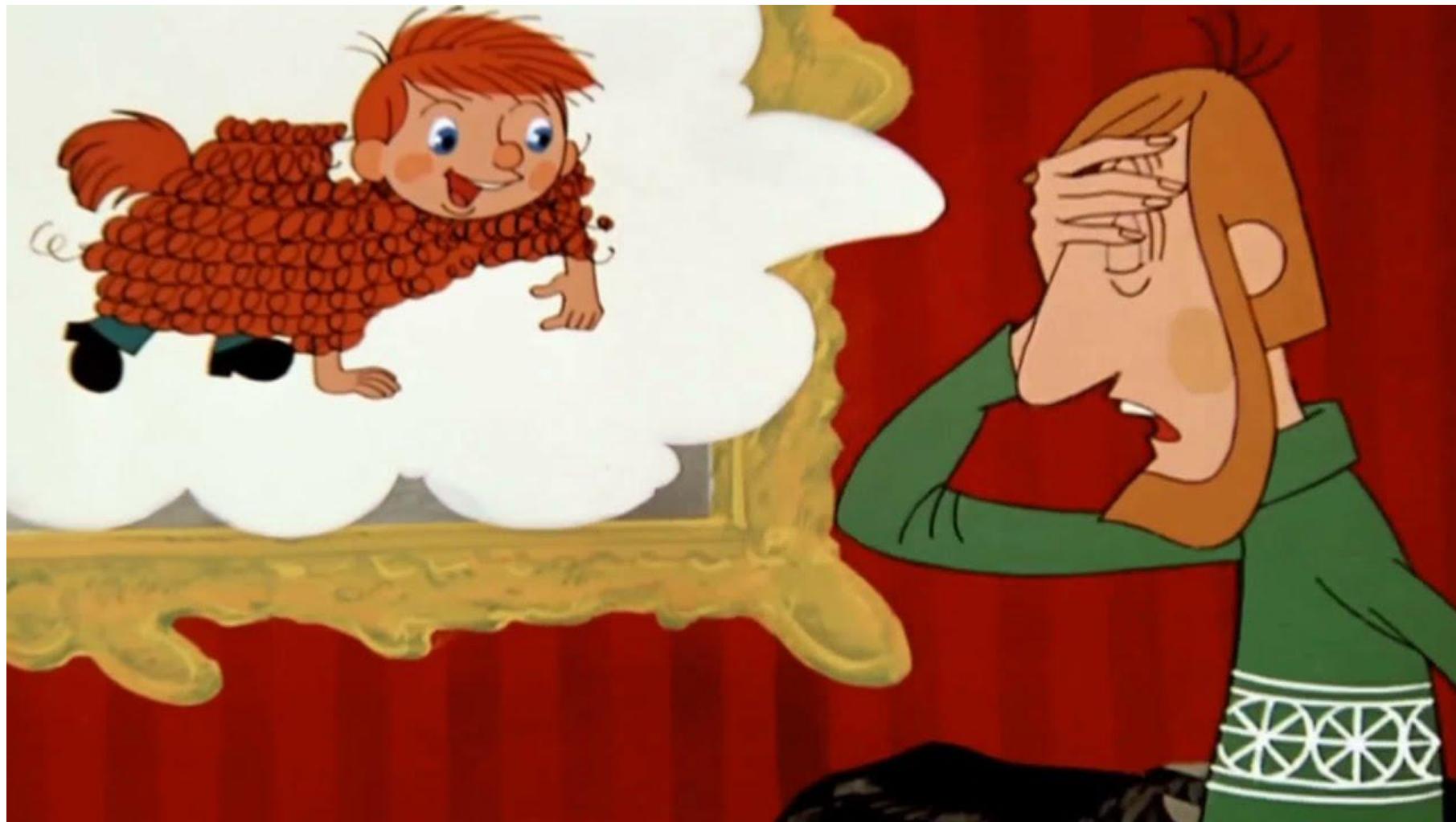
Как вносить исправления?

ID записи	Ошибка	Какое исправление сделано	Кто внес исправление	Дата
ISEE-1245	Неправильно указана широта 45° 71.345'	Исправлено на 45.71345	Сидоров И.И.	2010-08-05
ISEE-8354	Дата сбора 30 февраля	Удалено	Пахомов А.Е.	2013-12-25
ISEE-0507	Дубль записи ISEE-05077	Запись 05077 удалена	Боровиков Н.Н.	2015-05-10
ISEE-8932	Фамилия коллектора указана неверно	Исправлено с Пономарев на Понамарев	Волков А.А.	2017-03-18

Тщательное документирование
Сохранение исходных данных (с ошибками)

Зачем документировать исправления?





Как GBIF контролирует качество данных

Get data How-to Tools Community About dryomys

Data quality requirements

Publishers play an essential role not simply in sharing datasets, but also in managing their quality, completeness and usefulness and ensuring their integration and value within GBIF's global knowledge base.



Male baya weaver (*Ploceus philippinus*) waits outside as a female inspects the nest he built from inside. Females baya weavers select mates on the basis of the quality of nest construction. Photo by Malay Mehta via iNaturalist. Research-grade observations, licensed under CC BY-NC 4.0.

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



< Occurrences 1

TABLE GALLERY MAP TAXONOMY METRICS DOWNLOAD

Scientific name	Country or area	Coordinates	Month & year	Occurrence status	Basis of record	Dataset	Kingdom
<i>Mesophyllum lichenoides</i> (J.Ellis) Me Lemoi...	Spain		2022 January	Present	Preserved specimen	SANT-Algae	Plantae
<i>Eptesicus fuscus</i> subsp. <i>fuscus</i> (Palisot de ...	United States of America	41.3N, 72.9W	2022 January	Present	Preserved specimen	Vertebrate Zoology Division - Mammalogy...	Animalia
<i>Araeopteron</i> Hampson, 1893	Chinese Taipei	25.0N, 121.5E	2022 January	Present	Human observation	Taiwan Moth Occurrence Data Collected Fr...	Animalia
<i>Cyanistes caeruleus</i> (Linnaeus, 1758)	Germany	52.3N, 8.9E	2022 January	Present	Human observation	naturgucker	Animalia
<i>Croton</i> L.	Brazil	9.7S, 55.5W	2022 January	Present	Preserved specimen	CNMT - Herbário Centro Norte Mato Grosse...	Plantae
• <i>platystele ovalifolia</i> (focke) garay & dunst.	Brazil	1.1S, 48.4W	2022 January	Present	Preserved specimen	IAN herbarium - Embrapa Amazônia Orient...	Plantae
<i>Nucella lapillus</i> (Linnaeus, 1758)	France	47.1N, 2.2W	2022 January	Present	Human observation	naturgucker	Animalia
<i>Gryllodes sigillatus</i> (Walker, 1869)	United States of America	21.3N, 158.0W	2022 January	Present	Preserved specimen	University of Hawaii Insect Museum	Animalia
<i>Geastrum schmidelli</i> Vittad.	Poland	52.6N, 21.3E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Exidia saccharina</i> Fr.	Poland	53.2N, 20.9E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Flammulina elastica</i> (Sacc.) Redhead & R.H...	Poland	53.0N, 20.7E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Flammulina elastica</i> (Sacc.) Redhead & R.H...	Poland	53.6N, 14.8E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Flammulina elastica</i> (Sacc.) Redhead & R.H...	Poland	53.5N, 14.9E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Calocera furcata</i> (Fr.) Fr.	Poland	53.5N, 14.9E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Holwaya mucida</i> (Schulzer) Korf & Abawi	Poland	53.4N, 14.6E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Phaeotremella foliacea</i> (Pers.) Wedin, J.C.Z...	Poland	51.8N, 15.6E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Exidia glandulosa</i> (Bull.) Fr.	Poland	53.5N, 14.9E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Trichaptum biforme</i> (Fr.) Ryvarden	Poland	53.0N, 21.0E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi
<i>Phylloporopsis nidulans</i> (Pers.) Singer	Poland	54.2N, 19.4E	2022 January	Present	Human observation	Register of protected and endangered fungi...	Fungi

https://www.gbif.org/occurrence/search?occurrence_status=present&q=



OCCURRENCE | 1 JANUARY 2022

Natrix natrix (Linnaeus, 1758)

Common Grass Snake In English Observed in Finland

Animalia > Chordata > Reptilia > Squamata > Colubridae > *Natrix*[DETAILS](#) [RELATED RECORDS](#)Species: *Natrix natrix* (Linnaeus, 1758)

Location: Finland

Basis of record: Human observation

Event ID: <http://tun.fi/KE.176/625a9f60d5de48b7f1c2a9eb#Gathering1>

Dataset: Löydös Open Finnish Observation Database

Publisher: Finnish Biodiversity Information Facility

Issues: Recorded date unlikely

**Event**

Term	Interpreted	Original	Remarks
Day	1		Recorded date unlikely
Month	1		Recorded date unlikely
Year	2022		Recorded date unlikely
Event date	2022-01-01T00:00:00	2022-01-01/2022-12-31	Recorded date unlikely
Event ID	http://tun.fi/KE....c2a9eb#Gathering1	http://tun.fi/KE....c2a9eb#Gathering1	Recorded date unlikely
		0 occurrences	

<https://www.gbif.org/occurrence/3752412362>



OCCURRENCE | 5 AUGUST 1878

Natrix natrix (Linnaeus, 1758)

Common Grass Snake In English Collected in Sweden

Animalia > Chordata > Reptilia > Squamata > Colubridae > *Natrix*

Occurrence

Term	Interpreted	Original	Remarks
Catalogue number	K 1804	K 1804	
Individual count	1	1	
Occurrence ID	urn:catalog:ZMO:Herp:K 1804	urn:catalog:ZMO:Herp:K 1804	
Occurrence status	PRESENT		Occurrence status inferred from individual count
Recorded by	Hansson	Hansson	

Location

Term	Interpreted	Original	Remarks
Continent	EUROPE	Europe	
Coordinate uncertainty in metres		0	Coordinate uncertainty metres invalid
Country or area	Sweden	Sweden	Country coordinate mismatch
Country code	SE		Country coordinate mismatch
Decimal latitude	0	0.0	Zero coordinate Country coordinate mismatch
Decimal longitude	0	0.0	Zero coordinate Country coordinate mismatch
Depth	0		Inferred
Depth accuracy	0		Inferred
Elevation	0		Inferred
Elevation accuracy	0		Inferred
Geodetic datum	WGS84		Geodetic datum assumed WGS84 Country coordinate mismatch
Locality	Strømstad	Strømstad	
Maximum depth in metres		0	Excluded
Maximum elevation in metres		0	Excluded
Minimum depth in metres		0	Excluded
Minimum elevation in metres		0	Excluded

<https://www.gbif.org/occurrence/199454806>



Occurrences 2

Search all fields

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

Occurrence status

Present

Licence

Scientific name

Vulpes zerda (Zimmermann, 1780)

Basis of record

Location

Administrative areas (gadm.org)

Coordinate uncertainty in metres

Year

Month

Dataset

Country or area

Continent

Issues and flags

Media type

Publisher

Institution code

Collection code

SEARCH OCCURRENCES | 593 RESULTS

Флаги на результатах поискового запроса

TABLE GALLERY MAP TAXONOMY METRICS [DOWNLOAD](#)

DOWNLOAD OPTIONS

	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
SIMPLE	✗	✓	✗	✓ (if available)	Tab-delimited CSV (for use in Excel, etc.) <small>?</small>	319 KB (71 KB zipped for download)
DARWIN CORE ARCHIVE	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV (for use in Excel, etc.) <small>?</small>	977 KB (216 KB zipped for download)
SPECIES LIST	✗	✓	✗	✗	Tab-delimited CSV (for use in Excel, etc.) <small>?</small>	

DOWNLOAD REPORT

Total: 593

Licence: CC BY-NC 4.0

Year range: 1800–2022

With year: 81 %

With coordinates: 56 %

With taxon match: 100 %

Known issues

A part of the GBIF processing is to flag occurrences that have suspicious fields

44 Occurrence status inferred from individual count | 25 Type status invalid | 17 Continent invalid | 9 Recorded date invalid |
5 Basis of record invalid | 3 Country derived from coordinates | 2 Country invalid | 2 Taxon match fuzzy | 1 Coordinate invalid |
1 Multimedia URI invalid

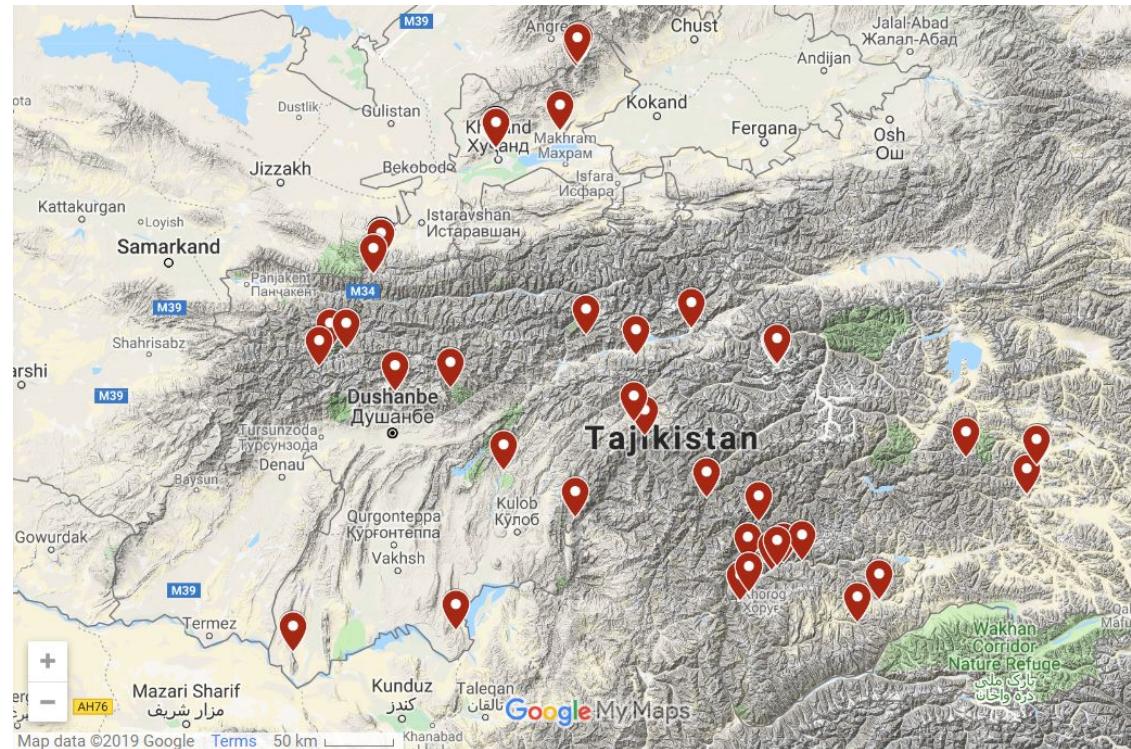


https://www.gbif.org/occurrence/download?taxon_key=5219310&occurrence_status=present

Практическое задание

Упражнение 1

Поиск и исправление
ошибок в данных
базовыми инструментами
электронных таблиц



Таксономический справочник GBIF

Get data How-to Tools Community About

NEWS | 25 SEPTEMBER 2019

GBIF introduces new version of the backbone taxonomy

Latest update adds 100,000 accepted species names



Sonora chub (*Gila dittaenia*) collected in Sonora, Mexico. Photo courtesy of University of Texas, Biodiversity Center, Ichthyology Collection (TNHC).

<https://www.gbif.org/news/2UfGq1L6iXbSu0ElamvVIH/gbif-introduces-new-version-of-the-backbone-taxonomy>

The Backbone taxonomy is actually a GBIF dataset. But not just any dataset, it is probably the most important dataset for GBIF.

Единая синтетическая классификация, которая преследует цель охватить все названия таксонов, с которыми имеет дело GBIF.

Не все используют одни и те же классификации или названия. Это приводит к значительным различиям в высших таксонах и большому количеству синонимов. GBIF Backbone призван объединить все эти названия и упорядочить их.



CHECKLIST DATASET | REGISTERED MARCH 2, 2011

GBIF Backbone Taxonomy

Published by [GBIF Secretariat](#)

[DATASET](#)[TAXONOMY](#)[CONSTITUENTS](#)[METRICS](#) [DOWNLOAD](#) [HOME PAGE](#)

6,783,300 RECORDS

78 CITATIONS

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



Publication date: November 26, 2021

Metadata last modified: December 9, 2021

Hosted by: [GBIF Secretariat](#)

Licence: CC BY 4.0

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



Для создания GBIF Backbone использовано [100 таксономических списков](#) по разным группам организмов. Для большинства таксонов приоритетным считается название из [Catalogue of Life](#).

<https://www.gbif.org/dataset/d7dddbf4-2cf0-4f39-9b2a-bb099caae36c>

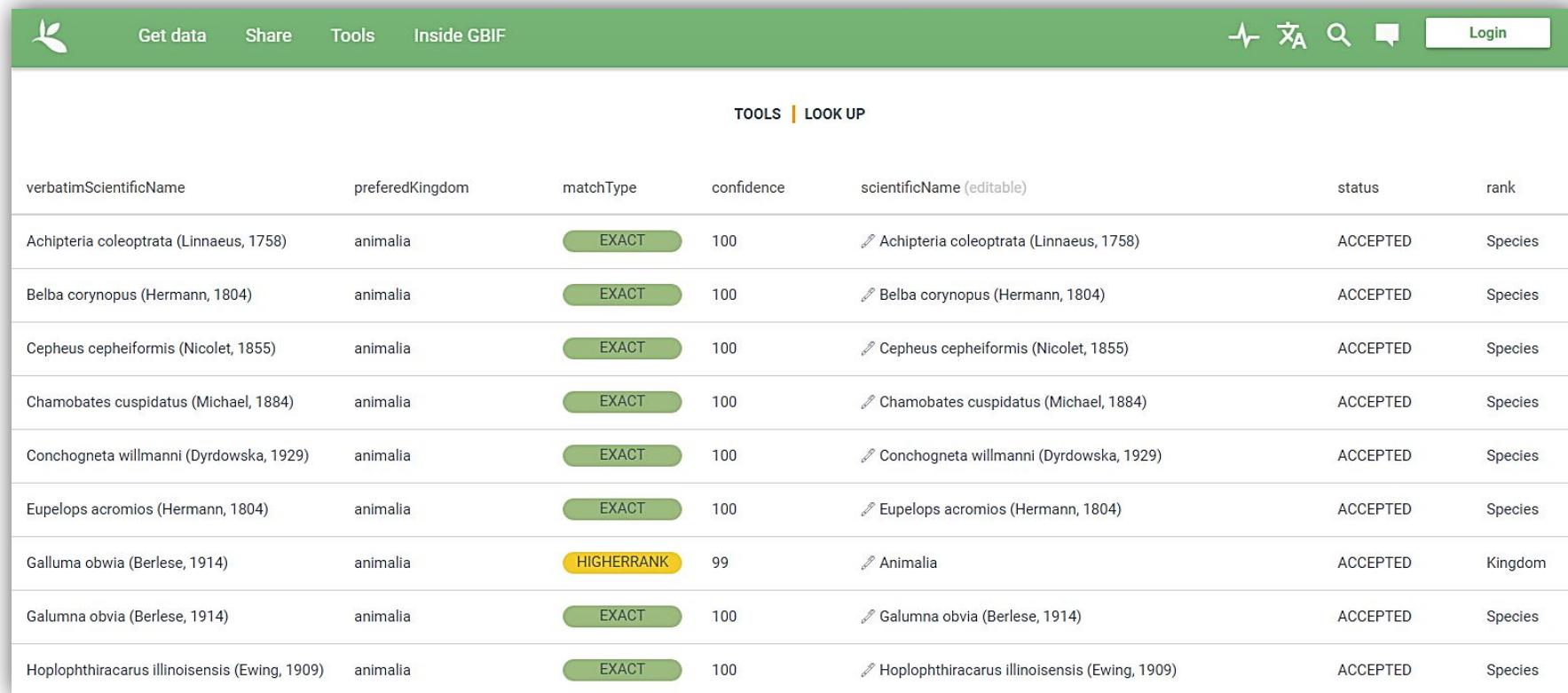
GBIF Backbone **содержит основные линнеевские ранги** от царства к виду плюс подвиды, разновидности и формы. Однако он **НЕ включает**:

- Неосновные линнеевские ранги (например, подсемейство, племя и т. д.),
- Гибридные формулы и сорта,
- Рабочие имена (кандидаты, имена-заполнители и нестабильные OTU).

Что делать, если названия нет в GBIF Backbone

- Оставьте название как есть, но добавьте таксоны более высокого ранга (алгоритм не распознает используемое научное название, но он все равно может найти таксон более высокого ранга).
- Если образец является типовым экземпляром, укажите его статус (dwc: typeStatus). Название будет интегрировано в GBIF Backbone при следующем обновлении.
- Проверить, нет ли ошибок в написании названия.

Проверка данных на соответствие базовой таксономии GBIF: поиск номенклатурных ошибок с помощью Species matching

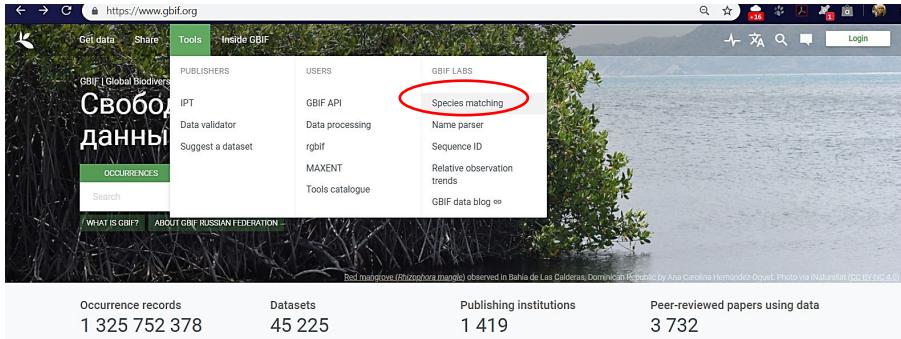


TOOLS	LOOK UP					
verbatimScientificName	preferedKingdom	matchType	confidence	scientificName (editable)	status	rank
<i>Achipteria coleoptrata</i> (Linnaeus, 1758)	animalia	EXACT	100	<i>Achipteria coleoptrata</i> (Linnaeus, 1758)	ACCEPTED	Species
<i>Belba corynopus</i> (Hermann, 1804)	animalia	EXACT	100	<i>Belba corynopus</i> (Hermann, 1804)	ACCEPTED	Species
<i>Cepheus cepheiformis</i> (Nicolet, 1855)	animalia	EXACT	100	<i>Cepheus cepheiformis</i> (Nicolet, 1855)	ACCEPTED	Species
<i>Chamobates cuspidatus</i> (Michael, 1884)	animalia	EXACT	100	<i>Chamobates cuspidatus</i> (Michael, 1884)	ACCEPTED	Species
<i>Conchogneta willmanni</i> (Dyrdowska, 1929)	animalia	EXACT	100	<i>Conchogneta willmanni</i> (Dyrdowska, 1929)	ACCEPTED	Species
<i>Eupelops acromios</i> (Hermann, 1804)	animalia	EXACT	100	<i>Eupelops acromios</i> (Hermann, 1804)	ACCEPTED	Species
<i>Galluma obvia</i> (Berlese, 1914)	animalia	HIGHERRANK	99	<i>Animalia</i>	ACCEPTED	Kingdom
<i>Galumna obvia</i> (Berlese, 1914)	animalia	EXACT	100	<i>Galumna obvia</i> (Berlese, 1914)	ACCEPTED	Species
<i>Hoplophthiracarus illinoiensis</i> (Ewing, 1909)	animalia	EXACT	100	<i>Hoplophthiracarus illinoiensis</i> (Ewing, 1909)	ACCEPTED	Species

Ошибки в названиях таксонов

Флаг	Значение	Пример
Taxon match fuzzy	Название таксона имеет небольшие отличия от названия этого таксона в GBIF	<i>Pelagodes antiquadrarius</i> и <i>Pelagodes antiquadraria</i>
Taxon Match higherrank	Не удалось найти таксон в GBIF, но нашелся таксон более высокого ранга	<i>Hylatomus pilatus</i> и <i>Hylatomus</i>
Taxon match none	Нет никаких совпадений	Flagellate

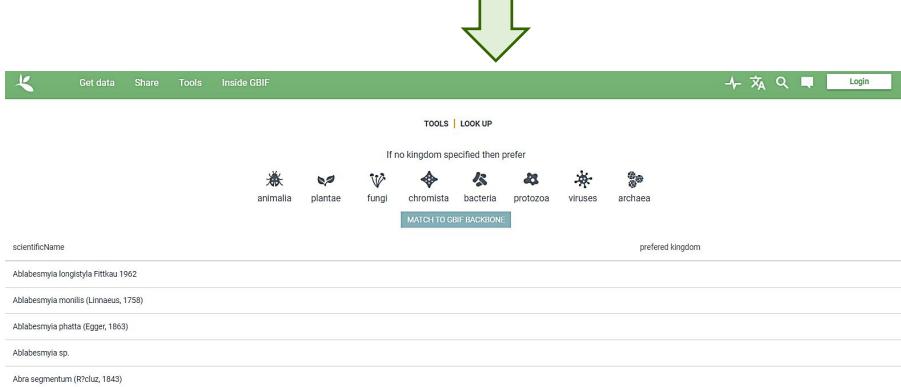
Species matching: инструкция по использованию (1/2)



Screenshot of the GBIF.org homepage. The 'Tools' menu is open, and the 'Species matching' option is circled in red. The main menu bar includes 'Get data', 'Share', 'Tools', and 'Inside GBIF'. Below the menu, there are sections for 'PUBLISHERS', 'USERS', and 'GBIF LABS'. The 'GBIF LABS' section contains links for 'Species matching', 'Name parser', 'Data processing', 'Data validator', 'Suggest a dataset', 'rbif', 'MAXENT', 'Tools catalogue', 'GBIF data blog', and 'Sequence ID', 'Relative observation trends'. Below the menu, there are statistics: 'Occurrence records 1 325 752 378', 'Datasets 45 225', 'Publishing institutions 1 419', and 'Peer-reviewed papers using data 3 732'. A large green arrow points downwards from the 'Species matching' link to the matching tool interface.

Инструмент Species matching находится в разделе Tools основного меню портала GBIF.org

Для сопоставления нужно скопировать названия таксонов в отдельную таблицу, назвать этот столбец scientificName и сохранить файл в формате CSV. Затем полученный файл нужно загрузить через диалоговое окно инструмента Species matching (<https://www.gbif.org/tools/species-lookup>) - выбрать его из папки (SELECT FILE), или просто перетащить (DROP HERE).

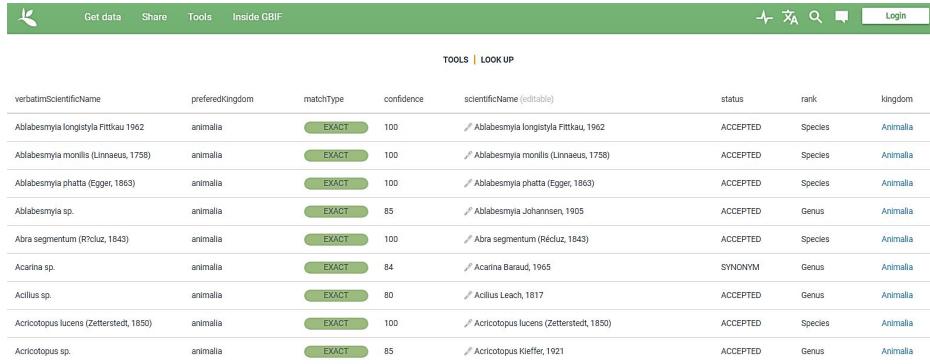


Screenshot of the 'Species matching' tool interface. The top navigation bar includes 'Get data', 'Share', 'Tools', 'Inside GBIF', and 'Login'. Below the navigation, there are 'TOOLS' and 'LOOK UP' buttons. A section titled 'If no kingdom specified then prefer' lists kingdom icons: animalia, plantae, fungi, chromista, bacteria, protozoa, viruses, and archaea. Below this is a 'MATCH TO GBIF BACKBONE' button. A table lists taxon names under 'scientificName': *Ablabesmyia longistyla* Fittkau 1902, *Ablabesmyia morilis* (Linnaeus, 1758), *Ablabesmyia phatta* (Egger, 1863), *Ablabesmyia* sp., and *Abra segmentum* (Rödl, 1843). A 'preferred kingdom' column is present but empty for these entries.

В открывшемся окне вы увидите список ваших таксонов. Если возможно, укажите Царство, к которому они относятся (кликнув соответствующую картинку), и нажмите кнопку MATCH TO GBIF BACKBONE

45

Species matching: инструкция по использованию (2/2)



The screenshot shows a search results page for species matching. At the top, there are navigation links: 'Get data', 'Share', 'Tools', 'Inside GBIF', and a search bar with a magnifying glass icon. Below the search bar are buttons for 'TOOLS' and 'LOOK UP'. The main content is a table with the following columns: 'verbatimScientificName', 'preferredKingdom', 'matchType', 'confidence', 'scientificName (editable)', 'status', 'rank', and 'kingdom'. The table lists several taxon records, each with a green 'EXACT' match type and 100 confidence. The 'status' column indicates whether the record is 'ACCEPTED' (Species, Genus, or Kingdom) or 'SYNONYM'. The 'rank' column shows the taxonomic rank (Species, Genus, or Animalia). The 'kingdom' column is consistently 'Animalia'. The table includes entries for *Ablabesmyia longistyla*, *Ablabesmyia monills*, *Ablabesmyia phutta*, *Ablabesmyia sp.*, *Abra segmentum*, *Acarina sp.*, *Acellus sp.*, *Acricotopus lucens*, and *Acricotopus sp.*.

verbatimScientificName	preferredKingdom	matchType	confidence	scientificName (editable)	status	rank	kingdom
<i>Ablabesmyia longistyla</i> Pittkau 1962	animalia	EXACT	100	<i>Ablabesmyia longistyla</i> Pittkau, 1962	ACCEPTED	Species	Animalia
<i>Ablabesmyia monills</i> (Linnaeus, 1758)	animalia	EXACT	100	<i>Ablabesmyia monills</i> (Linnaeus, 1758)	ACCEPTED	Species	Animalia
<i>Ablabesmyia phutta</i> (Egger, 1863)	animalia	EXACT	100	<i>Ablabesmyia phutta</i> (Egger, 1863)	ACCEPTED	Species	Animalia
<i>Ablabesmyia</i> sp.	animalia	EXACT	85	<i>Ablabesmyia</i> Johannsen, 1905	ACCEPTED	Genus	Animalia
<i>Abra segmentum</i> (Ricluz, 1843)	animalia	EXACT	100	<i>Abra segmentum</i> (Ricluz, 1843)	ACCEPTED	Species	Animalia
<i>Acarina</i> sp.	animalia	EXACT	84	<i>Acarina</i> Barraud, 1965	SYNONYM	Genus	Animalia
<i>Acellus</i> sp.	animalia	EXACT	80	<i>Acellus</i> Leach, 1817	ACCEPTED	Genus	Animalia
<i>Acricotopus lucens</i> (Zetterstedt, 1850)	animalia	EXACT	100	<i>Acricotopus lucens</i> (Zetterstedt, 1850)	ACCEPTED	Species	Animalia
<i>Acricotopus</i> sp.	animalia	EXACT	85	<i>Acricotopus</i> Kieffer, 1921	ACCEPTED	Genus	Animalia

Результат сопоставления доступен для просмотра в браузере, его можно сохранить в CSV файл, содержащий следующие поля: **occurrenceId** - идентификатор. Если вы его не указывали, поле будет пустым.

verbatimScientificName - предоставленные вами названия таксонов.

scientificName - название в соответствии с таксономической системой GBIF

key - идентификатор таксона в GBIF. При добавлении значения key к URL <https://www.gbif.org/species/> вы получите доступ к странице соответствующего таксона. Например, для *Bufo bufo* (Linnaeus, 1758) key=5217160, следовательно URL для страницы этого вида на портале GBIF <https://www.gbif.org/species/5217160>.

matchType - результат сопоставления.

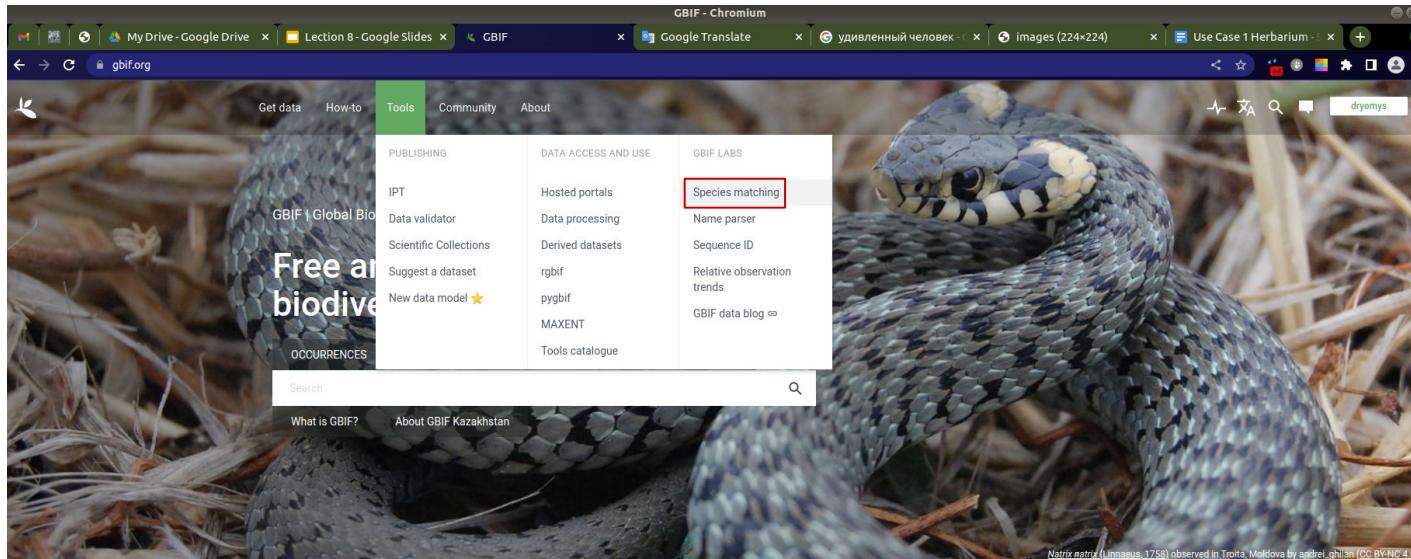
confidence - количественная характеристика степени совпадения

status - статус таксона в GBIF Backbone. Возможные значения: ACCEPTED - принятое в GBIF Backbone название, SYNONYM - синоним, DOUBTFUL - сомнительное.

rank - ранг таксона в GBIF Backbone на английском языке

Далее представлено положение таксона в иерархическом порядке, начиная с царства.

Демонстрация работы Species matching



<https://www.gbif.org/tools/species-lookup>

<https://www.gbif.org/tools/species-lookup>

Практическое задание

Упражнение 2

Проверка названий
таксонов
на соответствие
GBIF Backbone



Материалы для самостоятельного изучения

- Chapman AD (2005) Principles of Data Quality. Global Biodiversity Information Facility. <https://doi.org/10.15468/doc.jrgg-a190>
- Chapman AD & Wieczorek JR (2020) Georeferencing Best Practices. Copenhagen: GBIF Secretariat. <https://doi.org/10.15468/doc-gg7h-s853>
- Zermoglio PF, Chapman AD, Wieczorek JR, Luna MC & Bloom DA (2020) Georeferencing Quick Reference Guide. Copenhagen: GBIF Secretariat. <https://doi.org/10.35035/e09p-h128>
- Chapman AD & Grafton O (2008) Guide to Best Practices for Generalising Primary Species-Occurrence Data, version 1.0. Copenhagen: Global Biodiversity Information Facility, 27 pp. <http://www.gbif.org/resource/80512>
- Six questions answered about the GBIF Backbone Taxonomy <https://data-blog.gbif.org/post/gbif-backbone-taxonomy/>