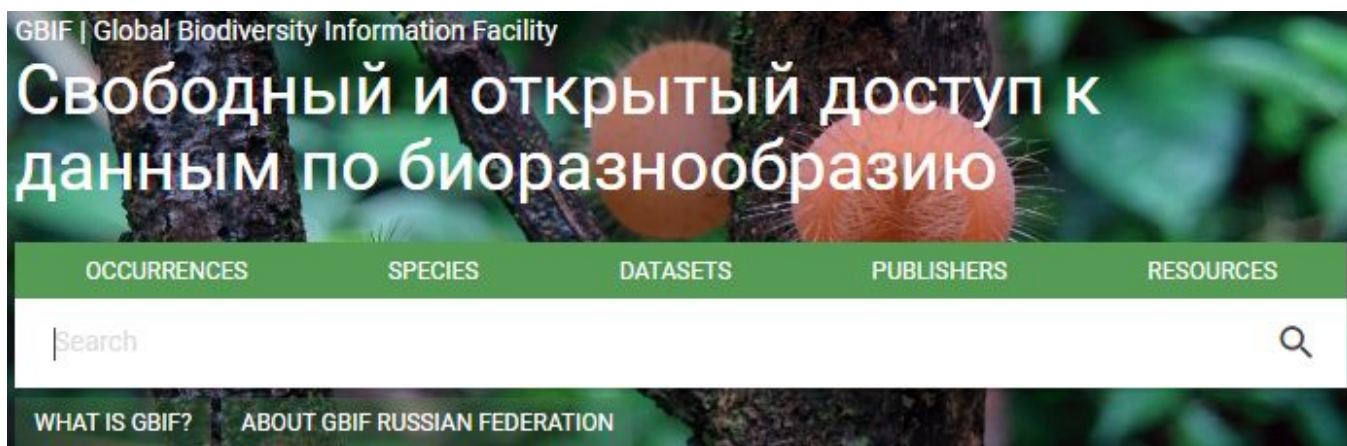


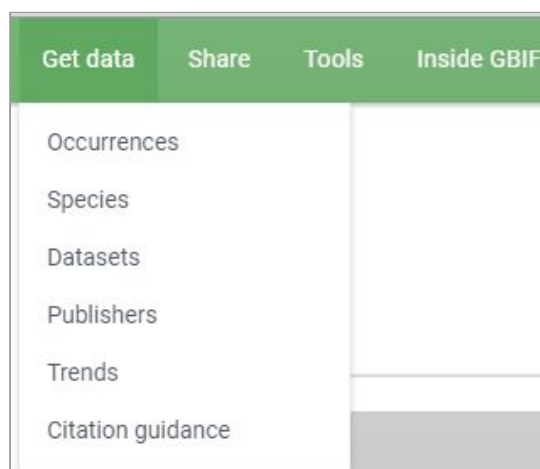
Как получить данные, опубликованные в GBIF

Большинство данных, доступных через портал [GBIF.org](https://www.gbif.org), представляют собой сведения об отдельных находках, для которых, как минимум, указано место и время находки, что было найдено, то есть видовое название организма или менее точное определение, а также кем эта находка (запись, наблюдение, ...) была сделана. На такого рода данных мы и будем учиться искать и получать нужную нам информацию.

Для начала нужно выбрать пункт **Occurrences** в основном меню поиска,



либо пункт **Occurrences** из раздела меню **Get data**, который доступен в любом месте портала GBIF. В появившейся таблице будут отображаться первые строки всего массива находок, опубликованных через GBIF. На сегодняшний день (2019-08-26) их более 1 млрд (1 337 051 293). Выберем находки какого-то одного вида, для этого введем его латинское название в поле **Scientific name**. Пусть это будет *Carabus nemoralis*. Можно задать дополнительные критерии поиска, например, год находки, страну или территорию и многие другие.



SEARCH OCCURRENCES | 1,337,051,293 RESULTS

Scientific name	Country or area	Coordinates
<i>Periparus ater</i> (Linnaeus, 1758)	Norway	59.5N, 5.3E
<i>Celenna festivariva</i> subsp. <i>formosensis</i> (Inou...	Taiwan	24.5N, 120.9E
<i>Cichladusa guttata</i> (Heuglin, 1862)	Kenya	3.0S, 38.0E
<i>Canis lupus</i> Linnaeus, 1758	Norway	60.3N, 12.2E
<i>Lynx lynx</i> (Linnaeus, 1758)	Norway	64.6N, 12.1E
<i>Gulo gulo</i> (Linnaeus, 1758)	Norway	67.8N, 16.1E

В отфильтрованной таблице возможно увидеть подробную информацию для отдельных находок, нажав на соответствующую запись. Кроме того во вкладке **Gallery** доступны фотографии объектов, во вкладке **Map** можно увидеть карту с географией находок, **Taxonomy** - таксономические сведения. **Metrics** - разнообразные графики, например, распределение находок по месяцам, по годам, по наборам данных и прочим параметрам, при этом вид графиков и необходимый параметр можно задавать вручную.

Occurrences
SEARCH OCCURRENCES | 9,503 RESULTS

Search all fields

Simple Advanced

License

Scientific name

Carabus nemoralis O.F.Müller, 1764

Basis of record

TABLE GALLERY MAP TAXONOMY METRICS

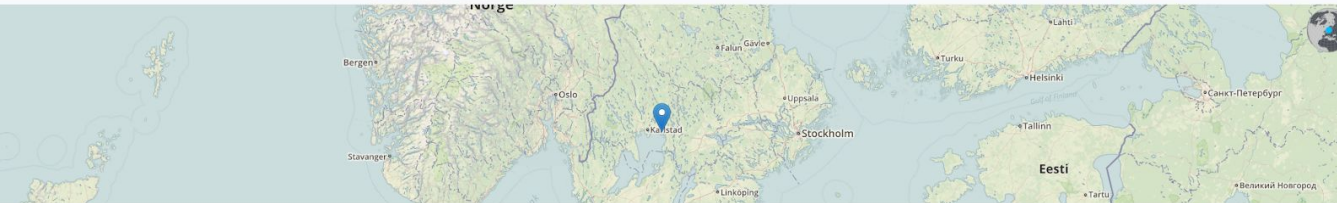
Scientific name	Country or area	Coordinates
<i>Carabus nemoralis</i> O.F.Müller, 1764	United States of America	48.0N, 122.9W
<i>Carabus nemoralis</i> O.F.Müller, 1764	Sweden	59.3N, 14.0E
<i>Carabus nemoralis</i> O.F.Müller, 1764	Norway	63.7N, 11.3E

OCCURRENCE | 4 JANUARY 2019
Modified 22 June 2019

Carabus nemoralis O.F.Müller, 1764

Hainlaufkäfer In German Observed in Sweden
 Animalia > Arthropoda > Insecta > Coleoptera > Carabidae > Carabus

Species: *Carabus nemoralis* O.F.Müller, 1764 Dataset: Artportalen (Swedish Species Observation System)
 Location: Europe > Sweden Publisher: ArtDatabanken
 Basis of record: Human observation



Record	Interpreted	Original	Remarks
Term			
Dataset ID	urn:lsid:swedishl...se:DataProvider:1	urn:lsid:swedishl...se:DataProvider:1	
Information withheld	More information can be obtained from the Data Provider.	More information can be obtained from the Data Provider.	
Institution code	ArtDatabanken	ArtDatabanken	

TABLE GALLERY MAP TAXONOMY METRICS


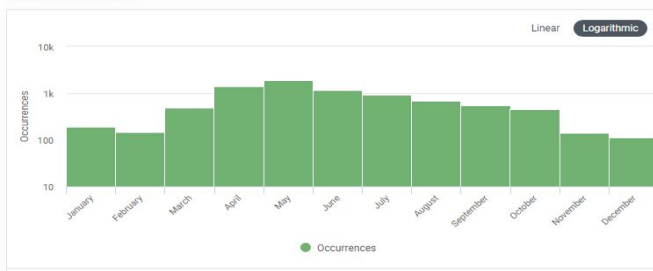


TABLE GALLERY MAP TAXONOMY METRICS

CURRENT FILTER Custom

Scientific name


OCCURRENCES PER MONTH



● Occurrences

1,366 other or unknown

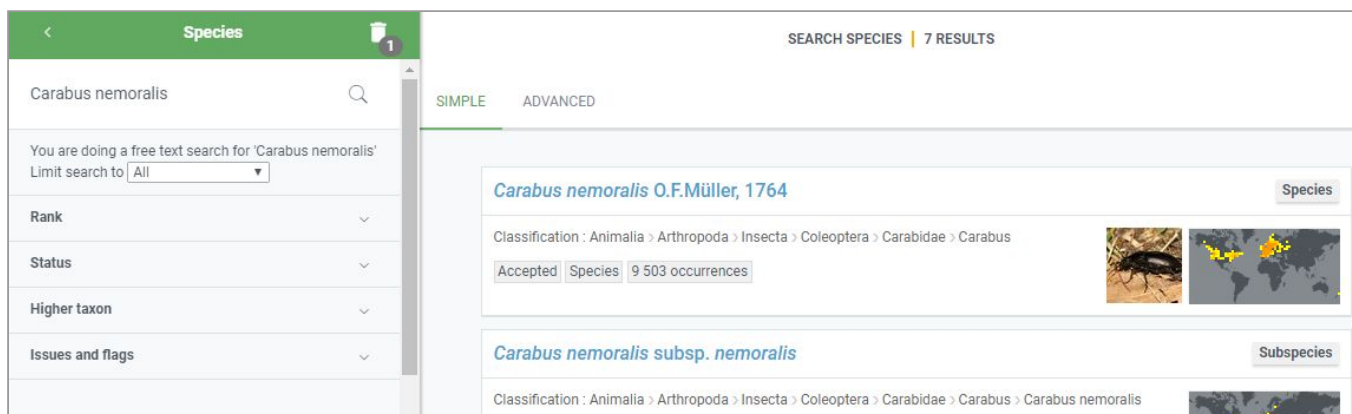
OCCURRENCES PER YEAR



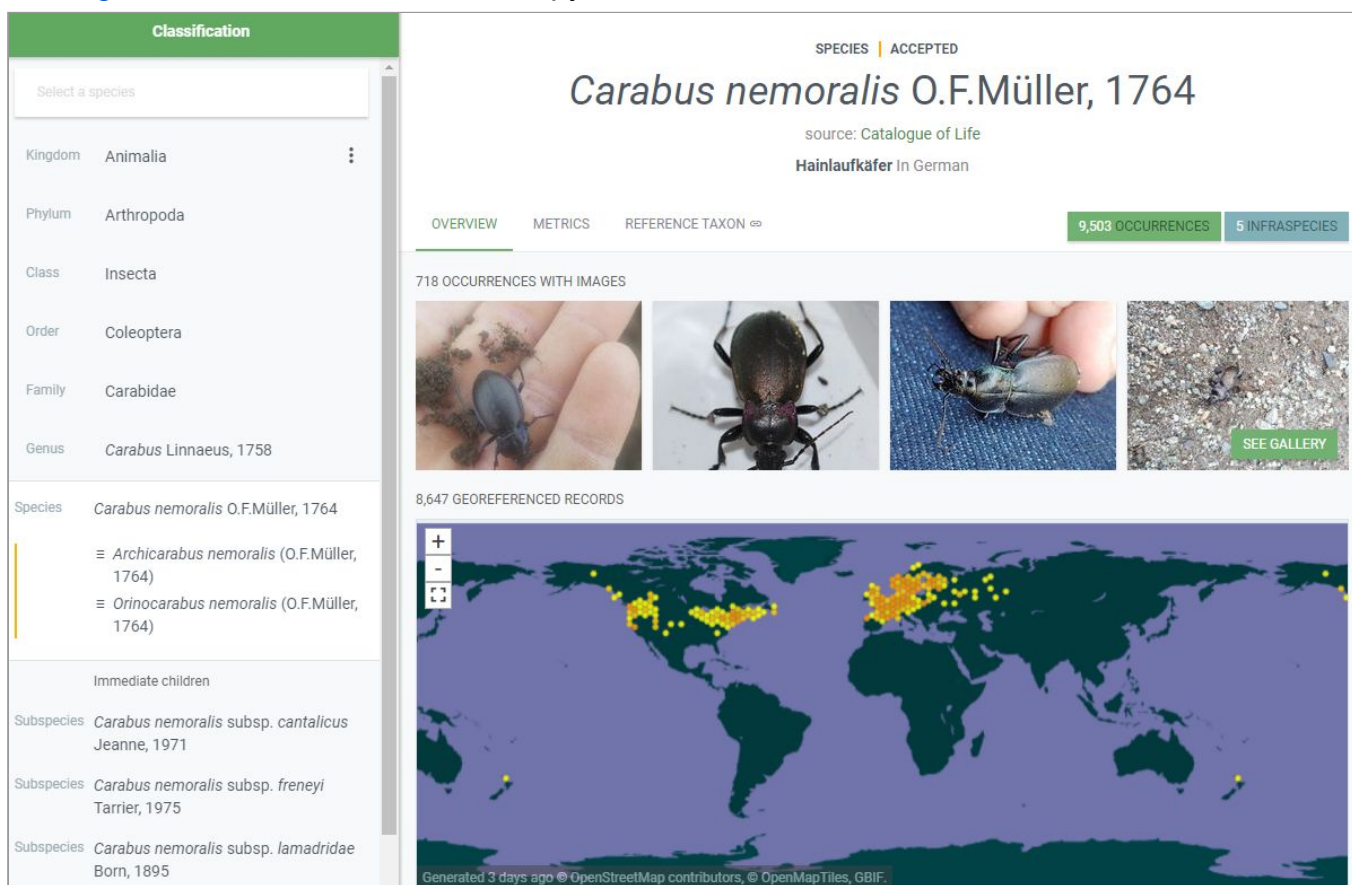
● Occurrences

1,111 other or unknown

Для находок представителей отдельных видов или более крупных таксонов такой же результат можно получить через раздел поиска **Species** (поиск по таксономической базе GBIF). В строке поиска нужно ввести название таксона, а затем из появившегося списка выбрать нужный вариант.



Таким образом можно получить информацию о любом таксоне (несмотря на то, что пункт поиска называется **Species**) из таксономической системы [GBIF Backbone](#). В настоящее время в ней содержится 5 858 142 записи (из которых 2 259 455 являются синонимами) и включает в себя другие крупные таксономические ресурсы, такие как [Catalogue of Life](#), [WoRMS](#) и многие другие.



На странице таксона можно найти сводку об информации, которая доступна через GBIF: карту находок, примеры фотографий, синонимы латинского названия и прочую таксономическую информацию, названия на разных языках и многое другое. Кроме того, можно использовать приведенную информацию при написании научной публикации, для чего в конце страницы приведена библиографическая ссылка на данную страницу.

В верхней правой части страницы можно найти число находок (на зеленом фоне), если нажать на эту ссылку, то попадем на страницу с таблицей находок для данного вида, точно такую же, что приведена на странице 2. Далее, если перейти на вкладку **Download**, то можно увидеть варианты получения сформированного вами набора данных и краткую сводку содержащейся в нем информации.

TABLE GALLERY MAP TAXONOMY METRICS **DOWNLOAD**

DOWNLOAD OPTIONS

	Raw data	Interpreted data	Multimedia	Coordinates	Format	Estimated data size
↓ SIMPLE	X	✓	X	✓ (if available)	Tab-delimited CSV ⓘ	4 MB (637 KB zipped for download)
↓ DARWIN CORE ARCHIVE	✓	✓	✓ (links)	✓ (if available)	Tab-delimited CSV ⓘ	11 MB (2 MB zipped for download)
↓ SPECIES LIST	X	✓	X	X	Tab-delimited CSV ⓘ	

DOWNLOAD REPORT

Total: 9 503
License: CC BY-NC 4.0
Year range: 1800–2019
With year: 88 %
With coordinates: 91 %
With taxon match: 100 %

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

527	Basin of record invalid	437	Recorded date mismatch	425	Country derived from coordinates	152	References uri invalid
63	Individual count invalid	58	Coordinate invalid	52	Taxon match higherrank	43	Coordinate precision invalid
						37	Recorded date invalid

Есть три варианта “комплектации” данных:

Simple - один файл с таблицей, которая содержит большую часть информации, относящуюся непосредственно к самим находкам. Расширение файла CSV.

Darwin Core Archive - архив содержащий не только таблицу с данными, относящимися к находкам, но и массу сопутствующей информации: описания наборов данных, из которых взяты эти находки, в формате XML вместе с библиографическими ссылками и лицензиями, относящимися к этим же наборам данных, в отдельных файлах; ссылки на фотографии и прочие медиа-файлы, связанные с находками; исходную неформализованную “дословную” информацию (verbatim) и прочее.

Species list - таксономический список в виде одной таблицы, которая в данном случае будет включать не только выбранный вид, но также его синонимы и подвиды.

В разделе сводки (**Download report**) можно найти сведения о числе записей, найденных для загрузки, временном интервале находок и некоторые другие. Особого внимания заслуживают разного рода несоответствия в данных, например, некорректная дата находки, страна не была указана в явном виде и была определена по координатам, не была указана система координат и прочие. Один или несколько вариантов таких неточностей, предусмотренных в системе GBIF, могут быть указаны для отдельных находок и набора данных в целом. Подобные недочёты не препятствуют публикации данных, но автоматически выявляются при индексации данных. Указание на них Вы можете найти в поле **issue** таблицы с находками. Подобная информация позволяет нам оценить качество данных, что очень важно при дальнейшей работе с ними.

После того, как Вы выберете нужный формат данных, внимательно прочитаете уведомление и нажмёте кнопку **UNDERSTOOD**, начнётся процесс сбора данных.

DOWNLOAD | 26 AUGUST 2019

Under processing

DOI 10.15468/dl.zca5sj Preparing CANCEL

FILTER APPLIED 26 AUGUST 2019 RERUN QUERY

The download has been started and is currently being processed.

Please expect up to 3 hours for the download to complete. Most downloads will complete within 15 minutes.

A notification email with a link to download the results will be sent to the following address once ready: max.carabus@gmail.com

Citation: GBIF.org (26 August 2019) GBIF Occurrence Download <https://doi.org/10.15468/dl.zca5sj>
License: Unspecified
Make sure to read the [data user agreement](#) and [citation guidelines](#).

Scientific name Carabus nemoralis O.F.Müller, 1764 API

На центральном сервере GBIF хранится только необходимый минимум сведений о находках, полностью все данные находятся на серверах организаций под управлением IPT (ПО для публикации данных), поэтому потребуется некоторое время для сбора всех необходимых данных и организации их в требуемый формат, обычно это занимает не более получаса, но часто еще быстрее. После формирования необходимого массива данных, на странице загрузки появится кнопка **Download** и некоторая информация о сформированном наборе данных: число записей, размер файла, список наборов данных, из которых взята информация по вашему запросу и прочее. Также присваивается DOI, используя который Вы сможете цитировать полученные данные.

DOWNLOAD | 26 AUGUST 2019

9,503 occurrences downloaded

DOI 10.15468/dl.zca5sj DOWNLOAD

FILTER APPLIED 26 AUGUST 2019 RERUN QUERY

Citation: GBIF.org (26 August 2019) GBIF Occurrence Download <https://doi.org/10.15468/dl.zca5sj>
License: CC BY-NC 4.0
File: 504 KB Simple
Involved datasets: 256
Make sure to read the [data user agreement](#) and [citation guidelines](#).

Scientific name Carabus nemoralis O.F.Müller, 1764 API

INCLUDES RECORDS FROM 256 DATASETS

Alterra (NL) - Entomofauna inventory in peat swamps	1
---	---

Также на ваш eMail придет письмо с прямой ссылкой на архив с данными.

- Inbox 237
- Snoozed
- Important
- Sent
- Drafts 100
- Spam 10
- Categories

Your GBIF data download is ready Inbox x

downloads@gbif.org
to me ▾

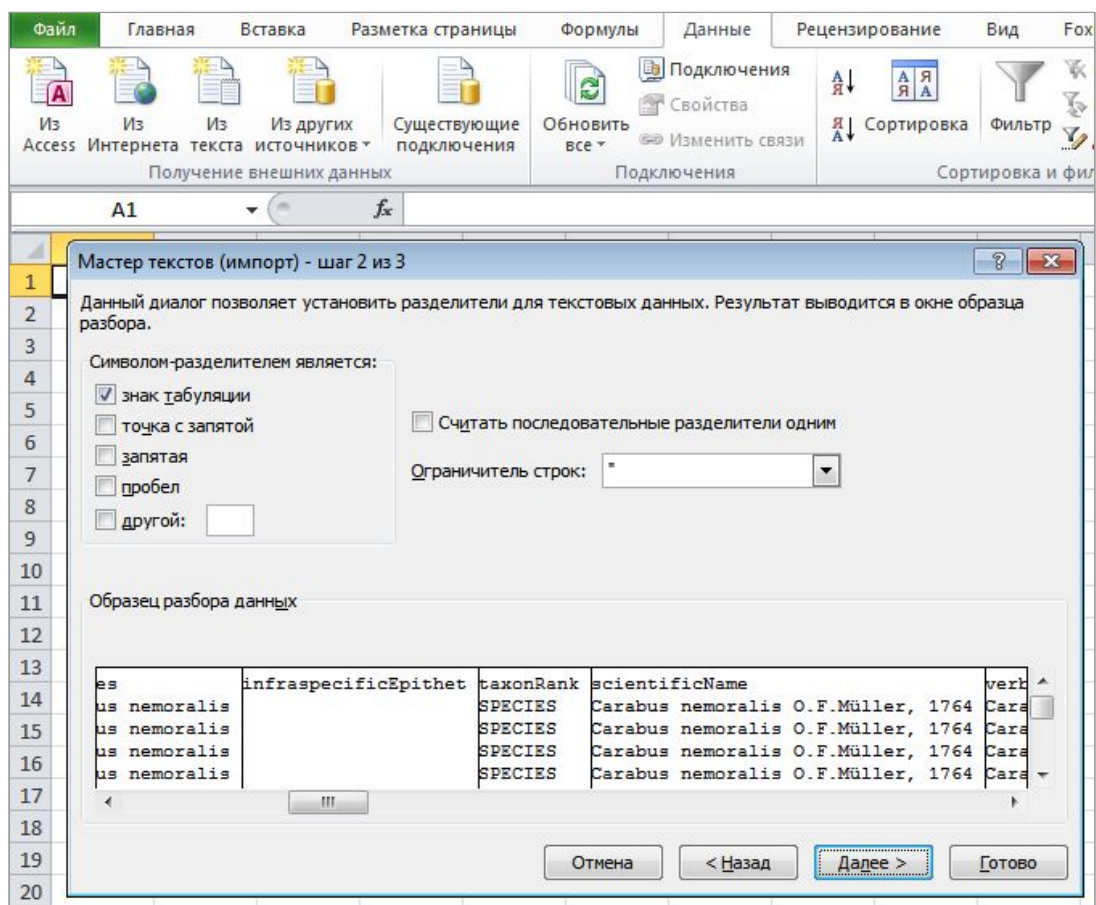
Hello max_carabus,

Your download is available at the following address:
<http://api.gbif.org/v1/occurrence/download/request/0005474-190813142620410.zip>

When using this dataset please use the following citation:
GBIF.org (26 August 2019) GBIF Occurrence Download <https://doi.org/10.15468/dl.zca5sj>

Ознакомление с данными лучше начать с варианта **Simple**. Несмотря на то, что табличные файлы, загружаемые из GBIF, имеют расширение CSV (значения разделенные, запятыми), на самом деле в этих файлах разделителем является знак табуляции, в связи с чем при открытии подобного файла могут возникнуть сложности. Табличный процессор Microsoft Excel по умолчанию настроен на файлы CSV с запятыми в качестве разделителей, поэтому если просто щелкнуть по файлу, или выбрать его через диалоговое окно **Открыть** непосредственно в программе Excel, то скорее всего такой файл откроется некорректно - большинство значений одной записи (строки) будет объединено в одну ячейку или несколько, если в тексте будут попадаться запятые, которые в данном случае на самом деле не будут являться разделителями значений. Можете убедиться в этом самостоятельно.

Для того, чтобы открыть такой файл в программе Microsoft Excel необходимо создать новый пустой файл, затем в разделе меню **Данные** выбрать инструмент **“Получить внешние данные Из текста”**, затем выбрать нужный файл и в появившемся диалоговом окне выбрать знак табуляции как символ-разделитель.



Также необходимо помнить, что кодировка файлов, загружаемых из GBIF - UTF-8.

При открытии CSV файлов в программе **LibreOffice Calc** подобных сложностей возникать не должно. Каким бы способом Вы не пытались открыть текстовый файл с табличными данными, будет появляться диалоговое окно преобразования файла. Вам нужно будет только корректно выбрать разделитель и кодировку.

