



Гербарийді цифрлау және GBIF IPT арқылы деректерді жариялау
Оцифровка гербария и публикация данных через GBIF IPT
Herbarium data digitization and publishing through GBIF IPT

Стандарты для обмена данными о биоразнообразии

Карагандинский университет имени академика Е.А. Букетова, 15-17 марта 2023 г.

План лекции

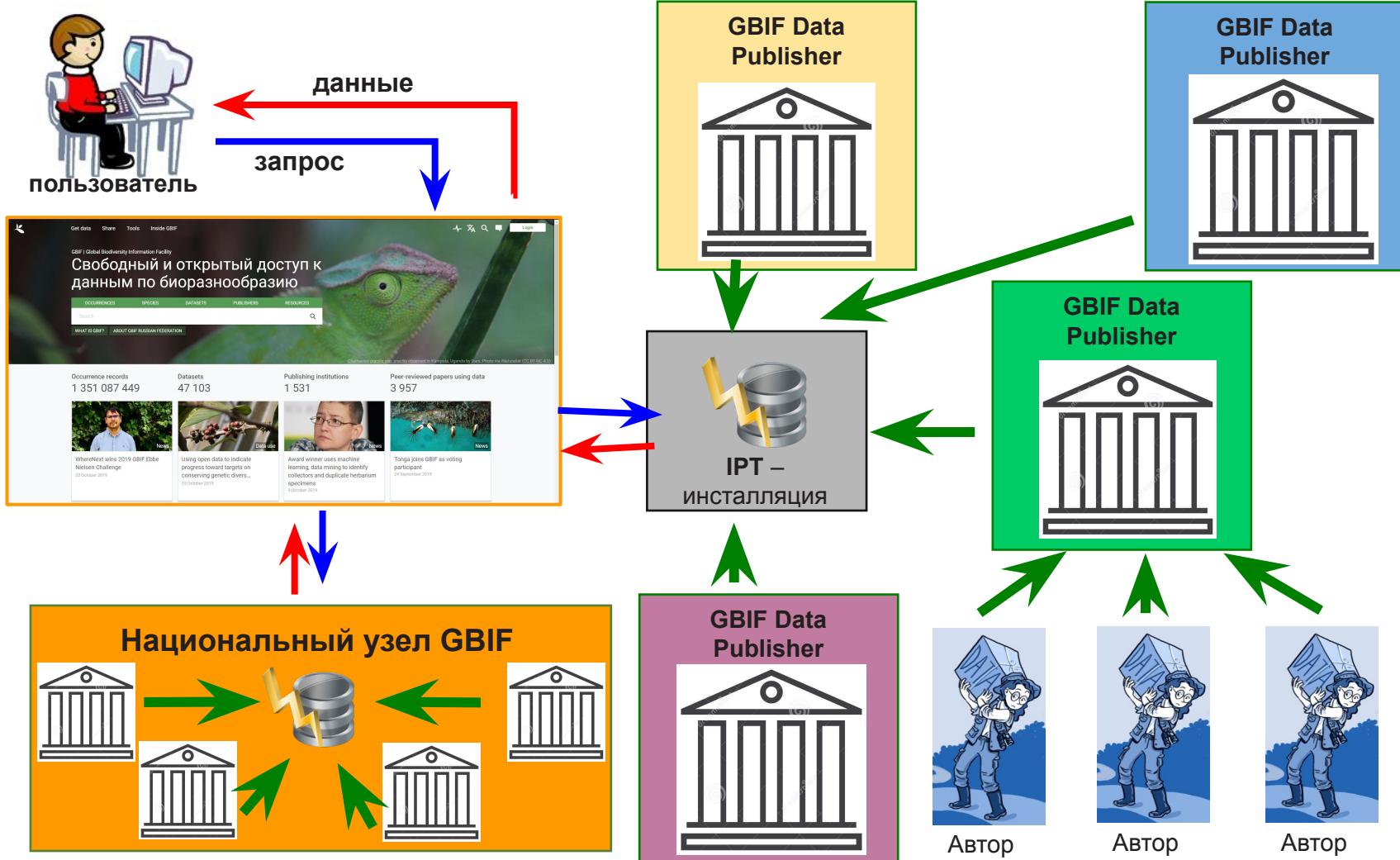
Стандарты в окружающем нас мире

Darwin Core

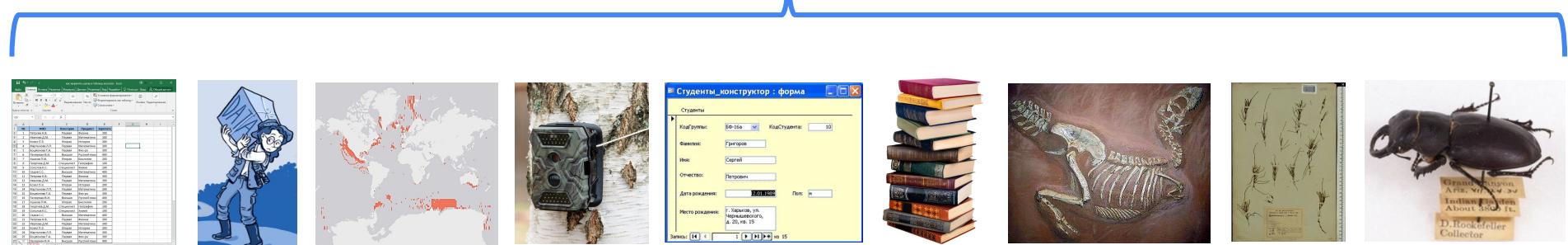
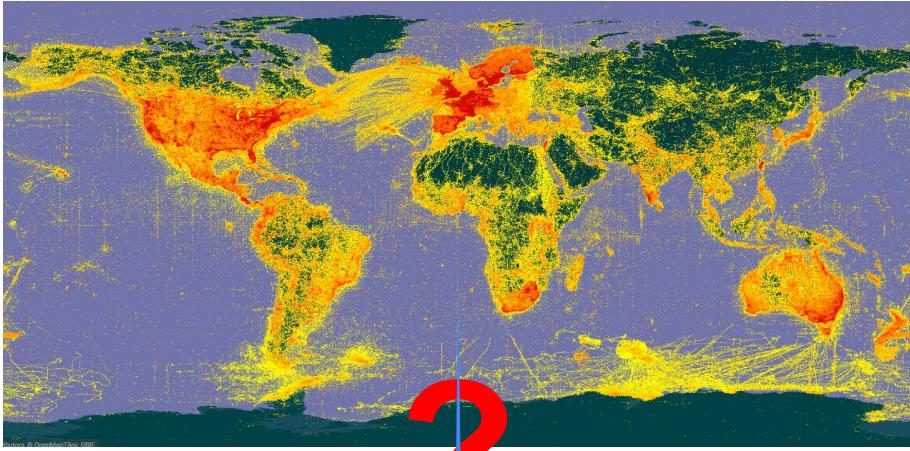
- Основные понятия и принципы использования
- Разделы Darwin Core
- Информационные ресурсы



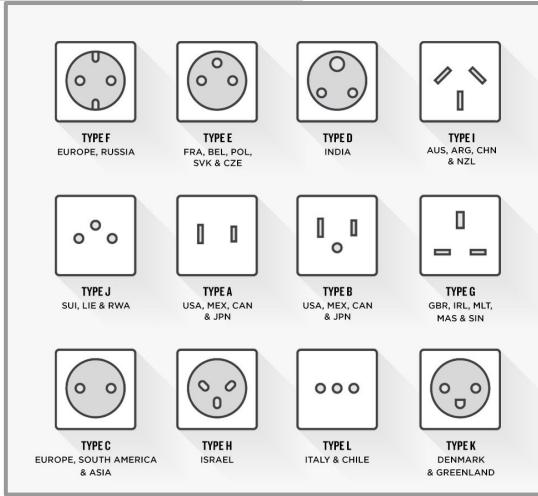
Как устроена сеть GBIF



Как объединить данные в глобальную систему?



ИСТОЧНИКИ ДАННЫХ



Повседневные стандарты

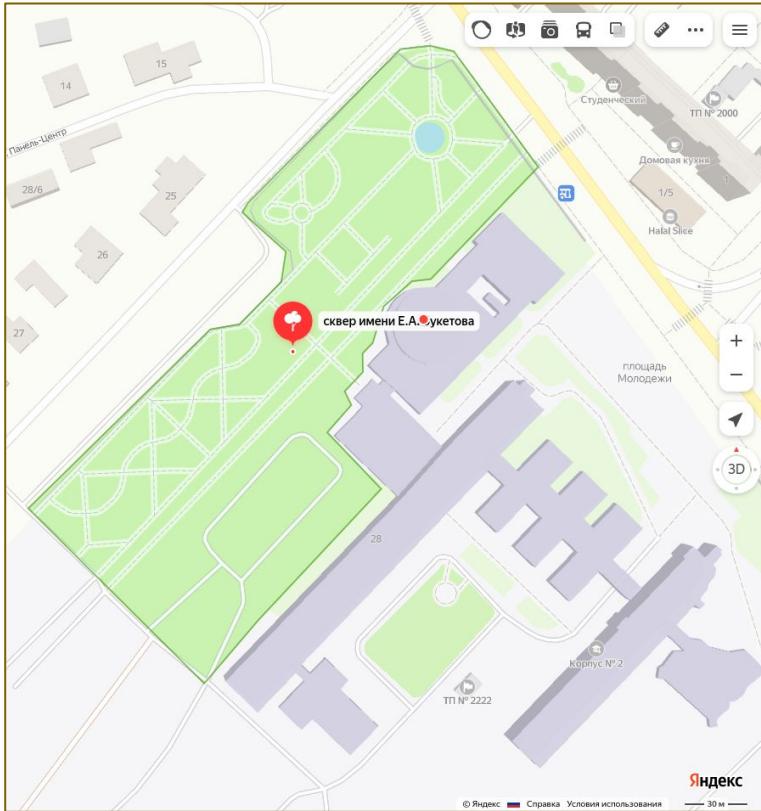
“Основная цель стандартов – создать основу для обмена информацией. Они должны обеспечивать ясность и способствовать коммуникации.”



Некоторые примеры стандартов, которые мы часто используем:

- Единицы измерения (Метрические)
 - Системы счисления
(Арабская, Римская)
 - Алфавиты
 - Языки
 - Смайлики
 - Почтовые индексы

Повседневные стандарты: пример



N49.77074, E73.12477

Формат: десятичные градусы

Цифры: арабские

Шрифт: Arial

Алфавит: латинский

International Organization for Standardization

Стандартизация во всех областях,
кроме электротехники и электроники



<https://www.iso.org/home.html>

ISO 3166-1: коды стран

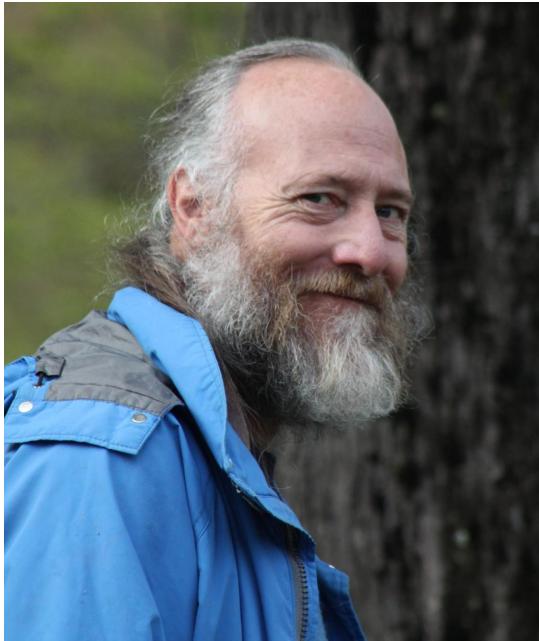
| ISO 3166-1 alpha-2 | ISO 3166-1 alpha-3 | ISO 3166-1 numeric |
|---|--|---|
| KZ | KAZ | 398 |
| Домены первого уровня https://buketov.edu.kz Номера на автомобилях | Машиночитаемые заграничные паспорта | Не зависит от алфавита. Используется в ООН |

Естественнонаучные стандарты

“Стандарты данных – это правила, по которым данные описываются и регистрируются. Для того, чтобы данные распространять, обмениваться ими и понимать их, мы должны стандартизировать как формат, так и смысл понятий.”
(Геологическая служба США)

- Язык стандарта экологических метаданных (EML),
- Описание мультимедийных материалов Audubon (Audubon Core),
- Global Глобальная сеть по биоразнообразию генома (GGBN),
- Проект передового опыта и стандартов об океанических данных (ODSBP),
- Darwin Core

Разработчики DwC



John Wieczorek
Convener

Core members

- Peter Desmet - Research Institute for Nature and Forest (INBO)
- Markus Döring - Global Biodiversity Information Facility (GBIF)
- Tim Robertson - Global Biodiversity Information Facility (GBIF)
- Steven Baskauf - Vanderbilt University Heard Libraries
- Paula Zermoglio - VertNet

Что такое Darwin Core (DwC)?

*“Список полей и их определения,
как они связаны с данными о биоразнообразии.”*



Biodiversity
Information
Standards
TDWG

TDWG [Home](#) [Terms](#) [Guides](#) ▾ [Namespace policy](#)

[GitHub](#)

Darwin Core

Darwin Core - это стандарт, поддерживаемый группой Darwin Core из TDWG. Он включает в себя словарь терминов (в разных ситуациях их также могут назвать свойствами, элементами, полями, столбцами, атрибутами или понятиями), предназначенных для облегчения обмена информацией о биологическом разнообразии, предоставляя идентификаторы, заголовки и определения.

В первую очередь, в основе Darwin Core находятся таксоны и их проявление в природе, подтвержденное наблюдениями, образцами, сборами и сопутствующей информацией.

Что такое Darwin Core (DwC)?

“Список полей и их определения,
как они связаны с данными о биоразнообразии.”

TDWG [Home](#) [Terms](#) [Guides](#) ▾ [Namespace policy](#)

[GitHub](#)

Darwin Core

Начало

- [Краткое руководство](#)
- Руководства по использованию: как использовать Darwin Core как [Simple Darwin Core](#), [text](#), [XML](#) или [RDF](#)
- [GitHub репозиторий](#): где поддерживается Darwin Core
- [Нормативные документ](#): CSV файл с полной версией истории терминов DwC
- [Файлы для распространения](#): удобные файлы для начала использования DwC

Стандарт Darwin Core -- dwc:basisOfRecord

Record-level

This category contains terms that are generic in that they might apply to any type of record in a dataset.



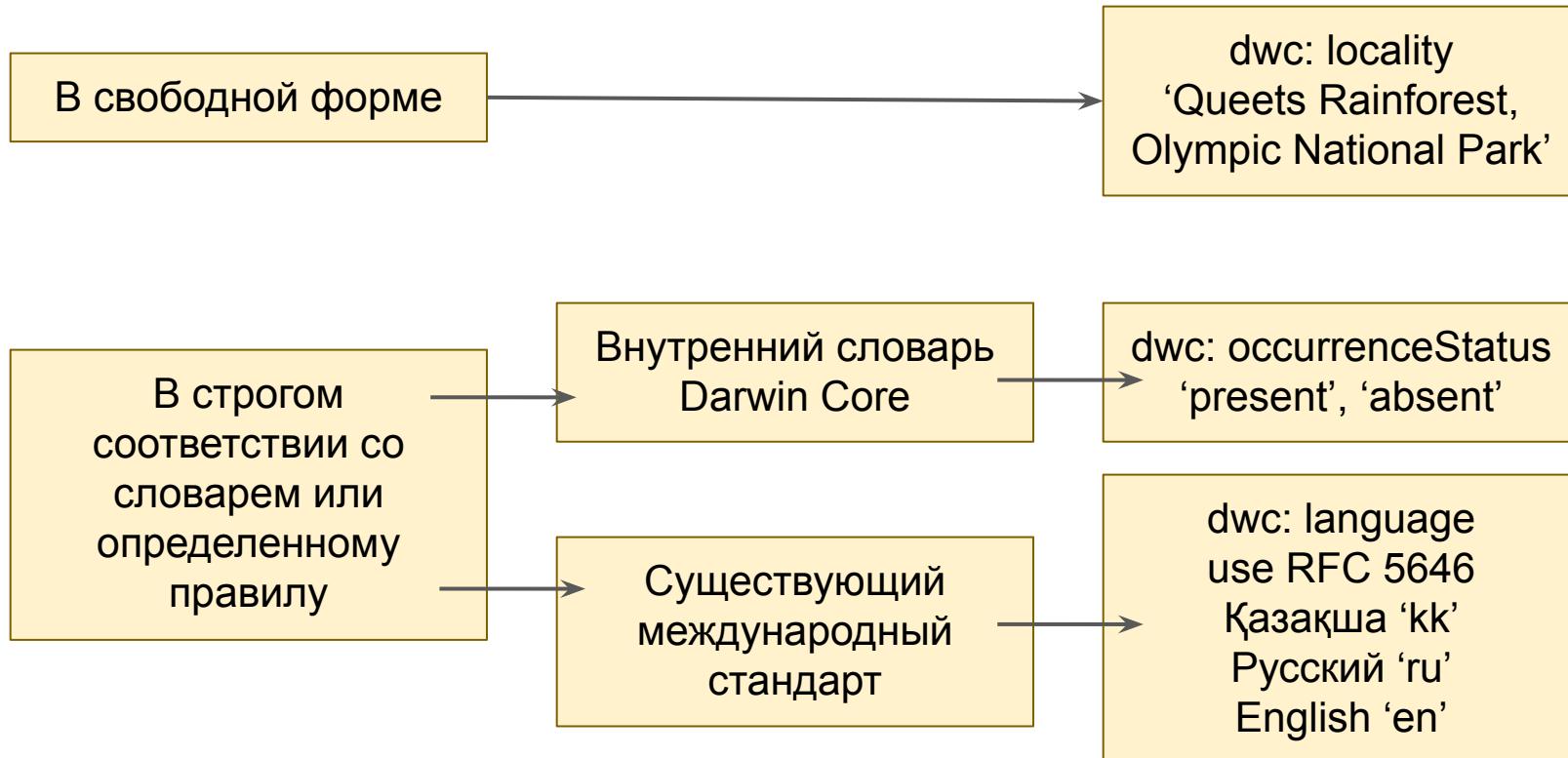
On this page

Record-level
Occurrence
Organism
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship
UseWithIRI
LivingSpecimen
PreservedSpecimen
FossilSpecimen
MaterialCitation
HumanObservation
MachineObservation
Cite Darwin Core

basisOfRecord

| | |
|------------|---|
| Identifier | http://rs.tdwg.org/dwc/terms/basisOfRecord |
| Definition | The specific nature of the data record. |
| Comments | Recommended best practice is to use the standard label of one of the Darwin Core classes. |
| Examples | PreservedSpecimen, FossilSpecimen, LivingSpecimen, MaterialSample, Event, HumanObservation, MachineObservation, Taxon, Occurrence, MaterialCitation |

Как вносить информацию в разные поля DwC



DwC: общие принципы заполнения полей

Термины с приставкой *verbatim*

dwc: verbatimEventDate
dwc: verbatimLocality
dwc: verbatimCoordinates
dwc: verbatimIdentification
dwc: verbatimIdentification
dwc: verbatimTaxonRank
и другие



как указано в
первоисточнике
(на языке
первоисточника)

Пример:

dwc: verbatimLocality
Дельта р. Урал в 10 км
сев. г. Атырау

dwc: locality
Delta of the Ural River,
10 km north of the city of
Atyrau

DwC: общие принципы заполнения полей

Термины с окончанием
remarks

dwc: occurrenceRemarks
dwc: organismRemarks
dwc: eventRemarks
dwc: locationRemarks
dwc: georeferenceRemarks
dwc: identificationRemarks
dwc: taxonRemarks

Для примечаний

Пример:

dwc: georeferenceRemarks
Assumed distance by road
(Hwy. 101)

DwC: общие принципы заполнения полей

Когда авторов (ссылок на литературу, названий институтов и т.д.)
несколько

Для разделения значений используется специальный символ |

Примеры:

| | |
|-------------------|---|
| dwc: recordedBy | José E. Crespo. Oliver P. Pearson Anita K. Pearson |
| dwc: recordedByID | https://orcid.org/0000-0002-1825-0097 https://orcid.org/0000-0002-1825-0098 |

DwC: что важно знать

dwc: basisOfRecord

Что стало основанием для записи в таблице?

Что является подтверждением встречи вида в природе?

Нужно взять подходящий вариант из словаря,
написание должно соответствовать словарю

Примеры

PreservedSpecimen - для коллекционных образцов

LivingSpecimen - для образцов из живых коллекций (бот. сад, зоопарк)

HumanObservation - для наблюдений и полевых сборов

MachineObservation - для данных, полученных при помощи автоматических регистраторов или других устройств.

DwC: что важно знать

Названия всех таксонов указывают в одном поле **dwc: scientificName** с той точностью, с которой удалось определить, **без пометок 'sp'**. Для указания **ранга таксона** используют **dwc: taxonRank**

| scientificName | taxonRank |
|--|-----------|
| Coleoptera | ORDER |
| Manis | GENUS |
| <i>Roptrocerus typographi</i> (Györfi, 1952) | SPECIES |

Ранги таксонов - см. словарь!

DwC: что важно знать

Географические координаты указывают в десятичных градуса, оптимально 5 знаков после запятой. Важно указывать не только сами координаты, но и систему, в которой они определены.

| | | |
|-----------------------|-------------------|----------|
| dwc: decimalLatitude | широта | 49.77074 |
| dwc: decimalLongitude | долгота | 73.12477 |
| dwc: geodeticDatum | система координат | WGS84 |

Если ваши координаты хранятся в формате ГГ ММ СС, их необходимо пересчитать в десятичные градуса (это очень легко).

DwC: что важно знать

Дата сбора (встречи, описания)

dwc: eventDate ISO 8601-1:2019 ГГГГ-ММ-ДД

| | |
|-----------------------|--|
| 2022-11-07 | 7 ноября 2022 г. |
| 2022-11-01/30 | 1-30 ноября 2022 г. |
| 2022-10-01/11-30 | с 1 октября по 30 ноября 2022 г. |
| 2021-01-01/2022-01-01 | с 1 января 2021 г. по 1 января 2022 г. |
| 2022 | в 2022 г. |
| 2022-11 | в ноябре 2022 г. |
| 2012/2022 | с 2012 по 2022 год (включительно) |

Разделы DwC

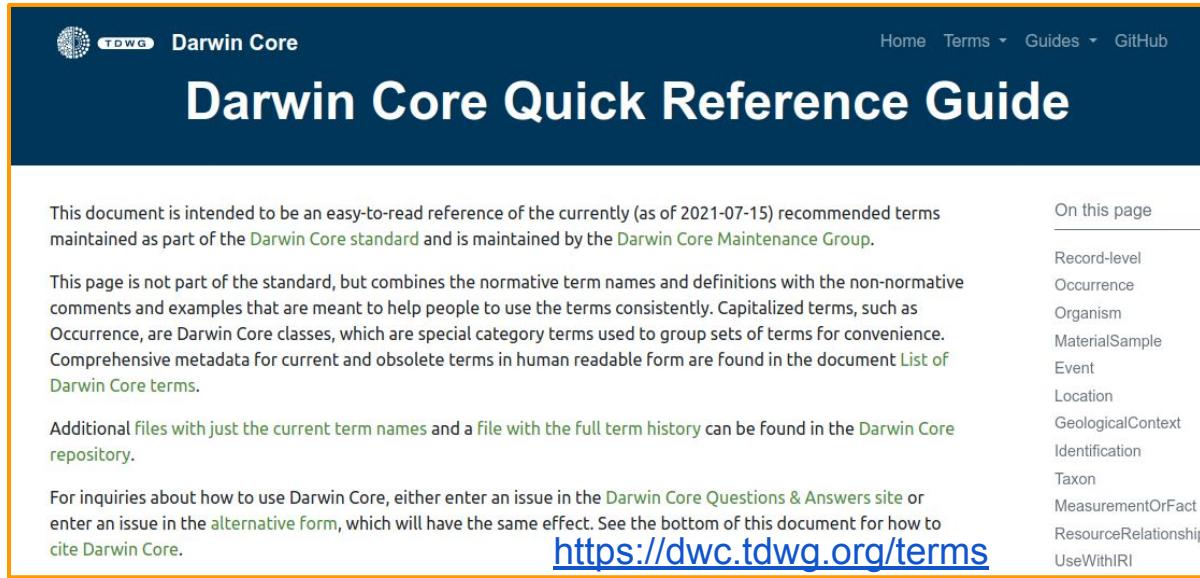
| | | |
|--------------------|--|---------------------------|
| Record-level Terms | Dublin Core terms, institutions, collections, nature of data record | Simple Darwin Core (flat) |
| Occurrence | evidence of species in nature, observers, behavior, associated media, references. | |
| Event | sampling protocols and methods, date, time, field notes | |
| Location | geography, locality descriptions, spatial data | |
| Identification | linkage between Taxon and Occurrence | |
| Taxon | scientific names, vernacular names, names usages, taxon concepts, and the relationships between them | |
| GeologicalContext | geologic time, chrono-stratigraphy, biostratigraphy, lithostratigraphy | |

Где используется DwC

The Global Biodiversity Information Facility (GBIF)
а также все национальные порталы стран-участниц GBIF

-  [The Ocean Biogeographic Information System \(OBIS\)](#) 107 805 309 occurrences
-  [The Atlas of Living Australia \(ALA\)](#) 112 452 049 occurrences
-  [VertNet](#) 171 collections from 12 countries and 52 additional collections (20 countries)
-  [Canadensys](#) 6 342 591 occurrences
-  [Encyclopedia of Life](#)
-  [Integrated Digitized Biocollections \(iDigBio\)](#) 134 737 435 specimens

Источники информации: на английском языке



The screenshot shows the homepage of the Darwin Core Quick Reference Guide. The header includes the TDWG logo and the text "Darwin Core". The main title is "Darwin Core Quick Reference Guide". Below the title, a paragraph explains the document's purpose: it is an easy-to-read reference of recommended terms from the Darwin Core standard, maintained by the Darwin Core Maintenance Group. The page is described as not being part of the standard but combining normative term names and definitions with non-normative comments and examples to help consistent use. It lists Darwin Core classes like Occurrence, MaterialSample, and Event. A link to the "List of Darwin Core terms" is provided. The page also links to files with current term names and a full history, and to the Darwin Core Questions & Answers site. A "Record-level" section on the right lists various term categories. A link to the "Darwin Core Hour" is present at the bottom.

This document is intended to be an easy-to-read reference of the currently (as of 2021-07-15) recommended terms maintained as part of the [Darwin Core standard](#) and is maintained by the [Darwin Core Maintenance Group](#).

This page is not part of the standard, but combines the normative term names and definitions with the non-normative comments and examples that are meant to help people to use the terms consistently. Capitalized terms, such as Occurrence, are Darwin Core classes, which are special category terms used to group sets of terms for convenience. Comprehensive metadata for current and obsolete terms in human readable form are found in the document [List of Darwin Core terms](#).

Additional files with just the current term names and a file with the full term history can be found in the [Darwin Core repository](#).

For inquiries about how to use Darwin Core, either enter an issue in the [Darwin Core Questions & Answers](#) site or enter an issue in the [alternative form](#), which will have the same effect. See the bottom of this document for how to cite Darwin Core.

<https://dwc.tdwg.org/terms>

On this page

Record-level
Occurrence
Organism
MaterialSample
Event
Location
GeologicalContext
Identification
Taxon
MeasurementOrFact
ResourceRelationship
UseWithIRI

Darwin Core Questions & Answers

<https://github.com/tdwg/dwc-qa>

GBIF community forum

<https://discourse.gbif.org>

Darwin Core Hour

<https://github.com/tdwg/dwc-qa/wiki/Webinars>



Источники информации: на русском языке

<http://gbif.ru> - русскоязычный информационный сайт о GBIF



Спецификация Darwin Core

На этой странице приводится описание терминов DwC на русском языке. Исходные описания терминов доступны [на сайте TDWG](#). Данная версия находится в разработке и будет обновляться. Предыдущая версия спецификации доступна [по ссылке](#).

Содержание терминов стандарта Darwin Core

Перевод и комментарии: Н. Иванова (*Институт математических проблем биологии РАН – филиал ИПМ им. М.В. Келдыша РАН*), М. Шашков (*Институт математических проблем биологии РАН – филиал ИПМ им. М.В. Келдыша РАН*). Апрель 2017 г. Оригинальная версия: <http://rs.tdwg.org/dwc/terms/>.

Разделы Darwin Core

[Record-level](#)
[Occurrence](#)
[Organism](#)
[MaterialSample](#)
[Event](#)
[Location](#)
[GeologicalContext](#)
[Identification](#)
[Taxon](#)
[MeasurementOrFact](#)
[ResourceRelationship](#)
[UseWithIRI](#)

- Перевод спецификации http://gbif.ru/DwC_spec/record-level (не обновляется с 2017 г.)
- Пособие по оцифровке Летописи природы <http://gbif.ru/guidebook> (Буйволов и др., 2019)

Заключение

Стандарты - это договоренности, используя которые, можно объединить данные из разных источников.

Для обобщения данных GBIF использует стандарт Darwin Core, который представляет собой набор полей в таблице и правила их заполнения.

Darwin Core постоянно развивается и дополняется. Для новых источников данных о биоразнообразии требуются новые термины.

Материалы для самостоятельного изучения

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, Giovanni R, et al. (2012) Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. <https://doi.org/10.1371/journal.pone.0029715>

BioData видео о стандартах <https://youtu.be/S02PJHPsRAs>

Guralnick R, Walls R, Jetz W. (2018) Humboldt Core – toward a standardized capture of biological inventories for biodiversity monitoring, modeling and assessment. Ecography, 41: 713-725. <https://doi.org/10.1111/ecog.02942>

Humboldt Core Task Group <https://github.com/tdwg/hc>