

«Подготовка данных для публикации в Глобальной информационной системе о биоразнообразии GBIF»

10 октября 2020 г., Екатеринбург

Лекция 1.3

Darwin Core: обменный стандарт для данных о биоразнообразии

Иван Чадин

Институт биологии Коми научного центра Уральского отделения Российской академии наук

Об эволюция баз данных о биологическом разнообразии

(одна из возможных интерпретаций)

Этап 1

База
данных

Отдельные (изолированные) базы данных на локальных компьютерах по отдельным коллекциям, таксономическим группам, территориям

База
данных

База
данных

База
данных

Этап 2

База
данных

База
данных

База
данных

База
данных

Стандарт

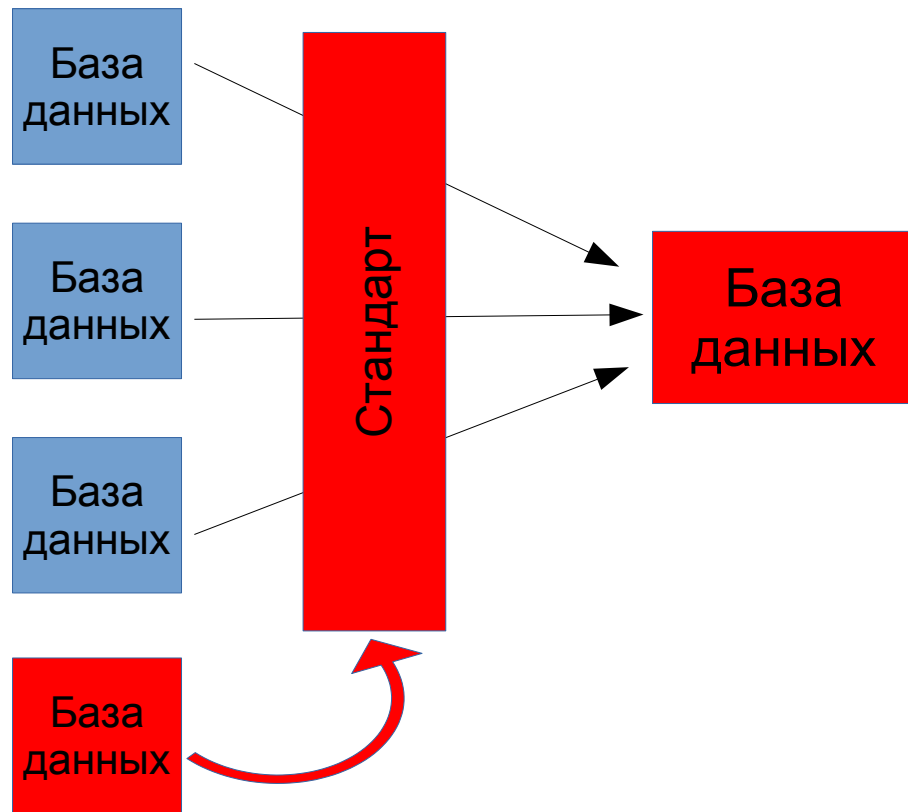
Осознание необходимости в разработке стандарта хранения данных для обеспечения возможности обмена данными

Появление различных инициатив по выработке стандартов и протоколов обмена данными

Появление TDWG (Taxonomic Database Working Group) ныне известной как некоммерческая организация и сообщество «Biodiversity Information Standards»

Этап 3

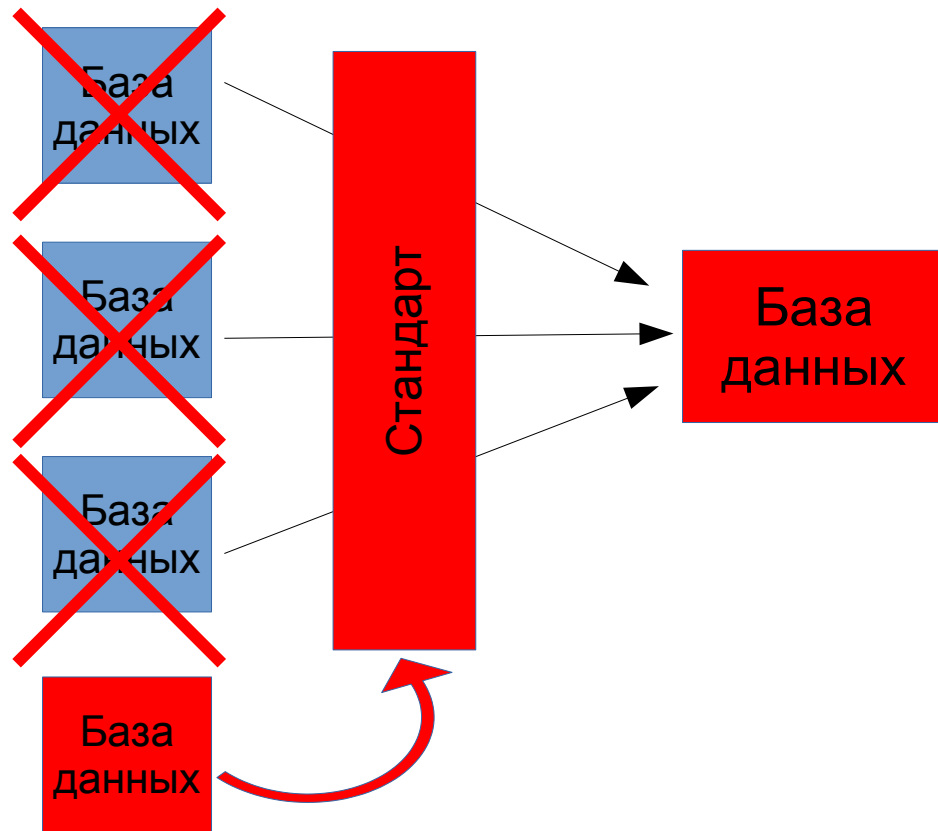
Попытки решить проблему выработки единого стандарта хранения и протокола обмена данными за счет принятия в качестве стандарта одной из существующих баз данных



Этап 3

Соблюдение стандартов и решение проблемы обмена данными проще всего решается прекращением поддержания отдельных баз данных после объединения всей информации в единую базу данных

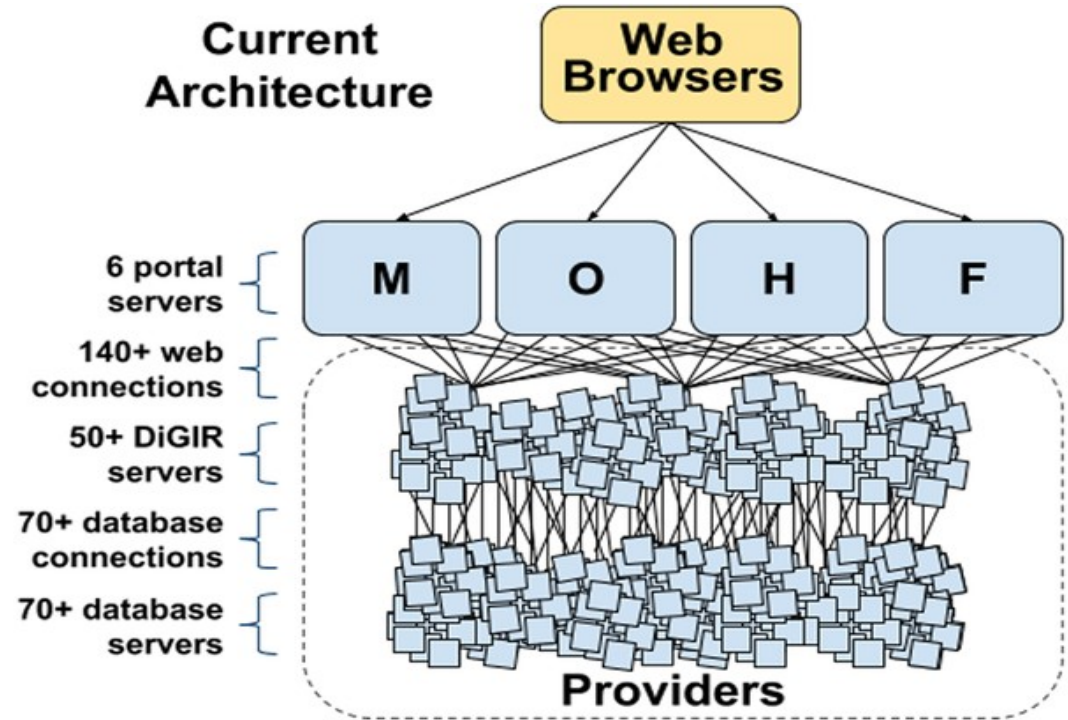
Не всех владельцев/кураторов баз данных такой подход приемлем



Этап 4

Принятие единого стандарта и протокола обмена данными.
Организация сложной системы, объединяющей отдельных поставщиков данных, зависящей от работоспособности отдельных узлов сети.

Организация сети обмена данными по позвоночным животным, некоторые элементы которой работают и в настоящее время. Использована трехуровневая система серверов, обменивающихся информацией по специальному протоколу DiGIR



Robertson T, Döring M, Guralnick R, Bloom D, Wiecek J, et al. (2014) The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLOS ONE 9(8): e102623. <https://doi.org/10.1371/journal.pone.0102623>

Этап 5.



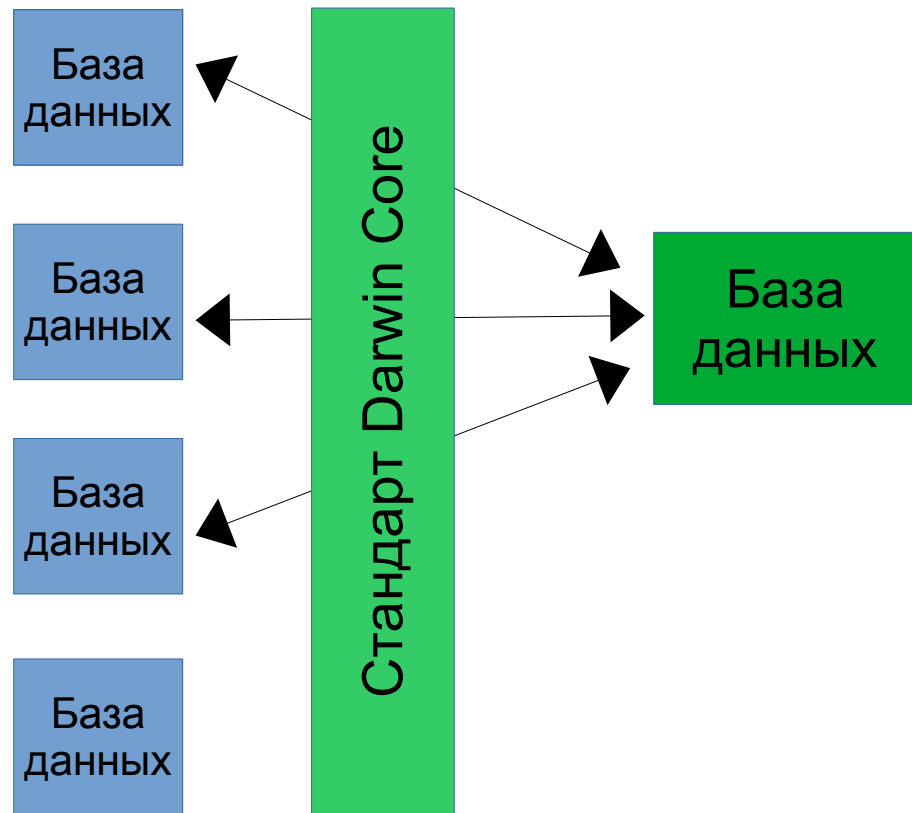
GBIF интегрирует информацию о тех базах данных, кураторы которых решили опубликовать свои данные.

Набор данных может быть отзован. Базы данных могут хранить информацию в произвольном формате.

Базы данных развиваются независимо от GBIF.

Основа для обмена данными – стандарт Darwin Core, выработанный профессиональным сообществом (TDWG).

GBIF для отдельной БД это аналог Яндекса для отдельного веб-сайта.



Darwin Core: An Evolving Community-Developed Biodiversity Data Standard

John Wieczorek, David Bloom , Robert Guralnick, Stan Blum, Markus Döring, Renato Giovanni, Tim Robertson, David Vieglais

Published: January 6, 2012 • <https://doi.org/10.1371/journal.pone.0029715>

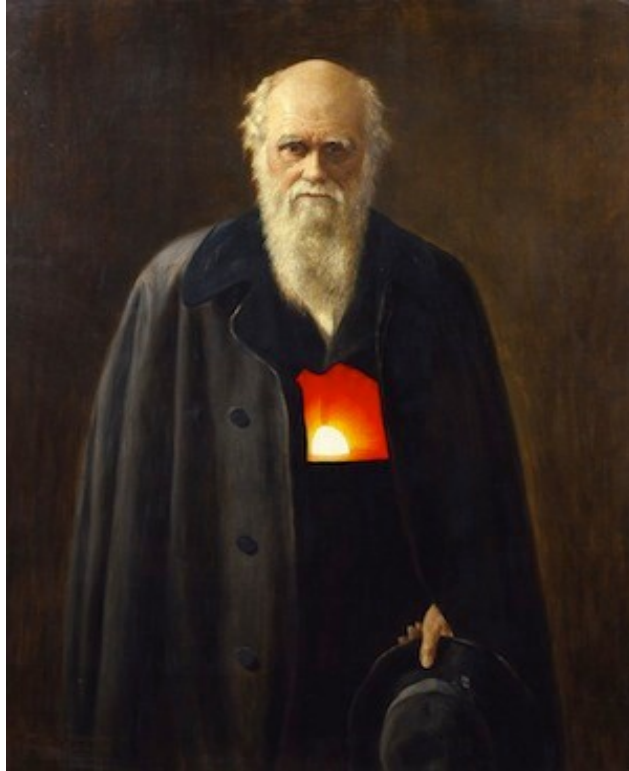
Darwin Core (DwC) — стандарт, созданный для обмена данными о биологическом разнообразии.

Разработана неформальной международной группой экспертов Biodiversity Information Standards (TDWG)

Позволяет публиковать данные в удобном человеко- и машино-читаемом формате, который может быть использован любым заинтересованным лицом.

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

Как, как вы сказали?
«Ядро Дарвина»???



innovation in metadata design, implementation & best practice

Register for DCMI Virtual 2020!

Dublin Core™ Metadata Initiative

[Home](#) [Specifications](#) ▾ [News](#) [Community](#) ▾ [Learning](#) ▾ [About](#) ▾ [Contact](#)



[Home](#) / [Specifications](#) / Dublin Core™

Dublin Core™

The original Dublin Core™ of thirteen (later fifteen) elements was first published in the [report of a workshop in 1995](#). In 1998, this was formalized in the Internet Engineering Task Force standard [RFC 5791](#), and [discussions began](#) about making it a standard of the (US) National Information Standards Organization (NISO). This led to the publication of ANSI/NISO Z39.85-2001 and the International Standards Organization Standard 15836-2003. The most recent updates of these standards are [RFC 5791 \(2010\)](#), [Z39-85-2012](#), and [ISO 15836-1:2017](#). Publication of a Part 2 to the ISO standard, covering several dozen properties and classes that have been added to DCMI namespaces since 1999, is expected in 2019.

Starting in 2002, DCMI grew into the role of "de facto" standards agency by maintaining its own, updated documentation for [DCMI Metadata Terms](#). The [DCMI Usage Board](#) currently serves as the maintenance agency for ISO 15836.

<https://dublincore.org/specifications/dublin-core/>

ФЕДЕРАЛЬНОЕ АГЕНТСТВО
ПО ТЕХНИЧЕСКОМУ РЕГУЛИРОВАНИЮ И МЕТРОЛОГИИ



НАЦИОНАЛЬНЫЙ
СТАНДАРТ
РОССИЙСКОЙ
ФЕДЕРАЦИИ

ГОСТ Р
7.0.10—
2010
(ИСО 15836:2003)

Система стандартов по информации,
библиотечному и издательскому делу

**НАБОР ЭЛЕМЕНТОВ МЕТАДААННЫХ
«ДУБЛИНСКОЕ ЯДРО»**

ISO 15836:2003
Information and documentation — The Dublin Core metadata element set
(MOD)

Издание официальное

5 Элементы метаданных

Т а б л и ц а 1 — Описание элементов «Дублинского ядра»

<u>Имя элемента</u>	<u>Описание элемента</u>
Title	Обозначение — Заглавие Определение — Наименование ресурса Комментарий — Как правило, Title — это наименование, под которым известен данный ресурс
Creator	Обозначение — Создатель Определение — Лицо, несущее первичную ответственность за создание ресурса Комментарий — Creator может быть человеком, организацией или службой. Как правило, этот элемент данных должен идентифицировать создателя
Subject	Обозначение — Предмет Определение — Предметное содержание ресурса Комментарий — Как правило, Subject выражается ключевыми словами, фразами или классификационными кодами, которые определяют тематику ресурса. Рекомендуется использовать контрольные словари или формализованные классификации ¹⁾
Description	Обозначение — Описание Определение — Обзор содержания ресурса Комментарий — Description может содержать аннотацию, оглавление, ссылку на графическое представление содержания, полнотекстовый реферат или аннотацию ²⁾ , а также другие данные
Publisher	Обозначение — Издатель Определение — Лицо, ответственное за предоставление ресурса пользователям Комментарий — Publisher может быть человеком, организацией или службой. Как правило, этот элемент данных должен идентифицировать издателя
Contributor	Обозначение — Соисполнитель Определение — Лицо, ответственное за вклад в содержание ресурса Комментарий — Contributor может быть человеком, организацией или службой. Как правило, этот элемент данных должен идентифицировать данное лицо
Date	Обозначение — Дата Определение — Дата некоторого события в жизненном цикле ресурса

Стандарт Darwin Core

Parts of the standard

The Darwin Core standard is comprised of one vocabulary, the Darwin Core vocabulary (<http://rs.tdwg.org/dwc/>), and six documents:

- [List of Darwin Core terms](#) - a complete list of all terms in namespaces currently used in the vocabulary
- [Simple Darwin Core Guide](#) - a guide for sharing information using the simplest methods
- [Darwin Core Text Guide](#) - a guide for sharing information using text files
- [Darwin Core XML Guide](#) - a guide for constructing application schemas using XML
- [Darwin Core RDF Guide](#) - a guide for encoding biodiversity data using the Resource Description Framework (RDF)
- [Darwin Core Namespace Policy](#) - an explanation of how URIs for Darwin Core terms should be constructed

An additional document, the [Quick Reference Guide](#) is maintained outside the standard and summarizes the currently recommended terms.

Стандарт Darwin Core

<https://www.tdwg.org/standards/dwc/>

Parts of the standard

The Darwin Core standard is comprised of one vocabulary, the Darwin Core vocabulary (<http://rs.tdwg.org/dwc/>), and six documents:

- [List of Darwin Core terms](#) - a complete list of all terms in namespaces currently used in the vocabulary
- [Simple Darwin Core Guide](#) - a guide for sharing information using the simplest methods
- [Darwin Core Text Guide](#) - a guide for sharing information using text files
- [Darwin Core XML Guide](#) - a guide for constructing application schemas using XML
- [Darwin Core RDF Guide](#) - a guide for encoding biodiversity data using the Resource Description Framework (RDF)
- [Darwin Core Namespace Policy](#) - an explanation of how URIs for Darwin Core terms should be constructed

An additional document, the [Quick Reference Guide](#) is maintained outside the standard and summarizes the currently recommended terms.

Стандарт Darwin Core

<https://www.tdwg.org/standards/dwc/>

Parts of the standard

The Darwin Core standard is comprised of one vocabulary, the Darwin Core vocabulary (<http://rs.tdwg.org/dwc/>), and six documents:

- [List of Darwin Core terms](#) - a complete list of all terms in namespaces currently used in the vocabulary
- [Simple Darwin Core Guide](#) - a guide for sharing information using the simplest methods
- [Darwin Core Text Guide](#) - a guide for sharing information using text files
- [Darwin Core XML Guide](#) - a guide for constructing application schemas using XML
- [Darwin Core RDF Guide](#) - a guide for encoding biodiversity data using the Resource Description Framework (RDF)
- [Darwin Core Namespace Policy](#) - an explanation of how URIs for Darwin Core terms should be constructed

An additional document, the [Quick Reference Guide](#) is maintained outside the standard and summarizes the currently recommended terms.

Darwin Core. Краткое руководство (quick reference guide)

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

Darwin Core quick reference guide

This document is intended to be an easy-to-read reference of the currently recommended terms maintained as part of the [Darwin Core standard](#). This page itself is not part of the standard. It draws on the term names and definitions from the normative part of the standard and combines them with comments and examples that are not normative, but that are meant to help people to use the terms consistently. Categories such as [Occurrence](#) and [Event](#) correspond to 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 a [list of terms document](#). [Files with lists of these terms](#) and [their full history](#) can be found in the [Darwin Core repository](#).

To cite the standard upon which this page is built, use the following:

Darwin Core Maintenance Group. 2020. List of Darwin Core terms. Biodiversity Information Standards (TDWG). <http://rs.tdwg.org/dwc/doc/list/>

To cite Darwin Core in general, use the peer-reviewed article on Darwin Core:

Wieczorek J, Bloom D, Guralnick R, Blum S, Döring M, 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>

Record-level

type	modified	language	license	rightsHolder	accessRights	bibliographicCitation	references	institutionID	collectionID
datasetID	institutionCode	collectionCode	datasetName	ownerInstitutionCode	basisOfRecord	informationWithheld			
dataGeneralizations	dynamicProperties								

[Record-level](#)

[Occurrence](#)

[Organism](#)

[MaterialSample](#)

[Event](#)

[Location](#)

[GeologicalContext](#)

[Identification](#)

[Taxon](#)

[MeasurementOrFact](#)

[ResourceRelationship](#)

[UseWithIRI](#)

[LivingSpecimen](#)

[PreservedSpecimen](#)

[FossilSpecimen](#)

[HumanObservation](#)

[MachineObservation](#)

Darwin Core. Краткое руководство (quick reference guide)

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

occurrenceID	catalogNumber	recordNumber	recordedBy	individualCount	organismQuantity	organismQuantityType	
sex	lifeStage	reproductiveCondition	behavior	establishmentMeans	occurrenceStatus	preparations	disposition
associatedMedia	associatedReferences	associatedSequences	associatedTaxa	otherCatalogNumbers	occurrenceRemarks		

Occurrence		Class
Identifier	http://rs.tdwg.org/dwc/terms/Occurrence	
Definition	An existence of an Organism (sensu http://rs.tdwg.org/dwc/terms/Organism) at a particular place at a particular time.	
Comments		
Examples	A wolf pack on the shore of Kluane Lake in 1988. A virus in a plant leaf in the New York Botanical Garden at 15:29 on 2014-10-23. A fungus in Central Park in the summer of 1929.	

occurrenceID		Property
Identifier	http://rs.tdwg.org/dwc/terms/occurrenceID	
Definition	An identifier for the Occurrence (as opposed to a particular digital record of the occurrence). In the absence of a persistent global unique identifier, construct one from a combination of identifiers in the record that will most closely make the occurrenceID globally unique.	
Comments	Recommended best practice is to use a persistent, globally unique identifier.	
Examples	http://arctos.database.museum/guid/MSB:Mamm:233627 , 000866d2-c177-4648-a200-ead4007051b9 , http://arctos.database.museum/guid/MSB:Mamm:233627	

Record-level

Occurrence

Organism

MaterialSample

Event

Location

GeologicalContext

Identification

Taxon

MeasurementOrFact

ResourceRelationship

UseWithIRI

LivingSpecimen

PreservedSpecimen

FossilSpecimen

HumanObservation

MachineObservation

Операция сопоставления названий столбцов в вашем наборе данных с терминами Darwin Core в программе IPT

Field Index

Record-level

MaterialSample

Occurrence

Organism

Event

Location

GeologicalContext

Identification

Taxon

Field Filters

Hide unmapped fields

Save

Delete

Back

materialSampleID

Save

Delete

Back

Occurrence

catalogNumber

Source Sample: 29138 | 8437 | 8461 | 47065 | 47066

Translation: Add

occurrenceRemarks

recordNumber

recordedBy

Source Sample: Sholeninova T.P. | Last Degteva S.V., Dubrovskiy Yu.A. | Degteva S.V.,

Translation: Add

individualCount

organismQuantity

organismQuantityType

sex

lifeStage

reproductiveCondition

behavior

establishmentMeans

occurrenceStatus

occurrenceID

InstitutionID

collectionCode

basisOfRecord

catalogNumber

recordedBy

associatedReferences

year

month

day

country

countryCode

locality

decimalLatitude

Термины Darwin Core (DwC)

Названия столбцов из вашего набора данных

Готовые шаблоны в формате Excel, в названиях столбцов которых использованы термины Darwin Core

Checklist Data

<https://github.com/gbif/ipt/wiki/checklistData#templates>

Occurrence Data

<https://github.com/gbif/ipt/wiki/occurrenceData#templates>

Sampling Event Data

<https://github.com/gbif/ipt/wiki/samplingEventData#templates>