

Федеральное государственное бюджетное учреждение  
“Приокско-Тerrasный государственный природный биосферный заповедник  
имени М.А. Заблoцкого”



# **ОЦИФРОВКА ДАННЫХ ЛЕТОПИСЕЙ ПРИРОДЫ И НАУЧНЫХ БИОЛОГИЧЕСКИХ КОЛЛЕКЦИЙ ОСОБО ОХРАНЯЕМЫХ ПРИРОДНЫХ ТЕРРИТОРИЙ**

Учебное пособие

Версия 1.0 (2019-11-30)

Данки  
2019

Буйолов Ю.А., Иванова Н.В., Шашков М.П. **Оцифровка данных Летописей природы и научных биологических коллекций особо охраняемых природных территорий.** Учебное пособие. ФГБУ Приокско-Террасный государственный природный биосферный заповедник, 2019.

Пособие подготовлено при финансовой поддержке GBIF FinBIF ([проект Russia2019\\_03](#))  
Раздел о проверке и верификации данных подготовлен на основе учебных материалов курсов повышения квалификации [BioDATA](#)

Авторы

**Юрий Анатольевич Буйолов** – к.б.н., заместитель директора по научной работе ФГБУ Приокско-Террасный государственный природный биосферный заповедник, пос. Данки, Московская область

**Иванова Наталья Владимировна** – к.б.н., научный сотрудник Института математических проблем биологии РАН – филиала ИПМ им. М.В. Келдыша, г. Пущино

**Шашков Максим Петрович** – старший научный сотрудник Института математических проблем биологии РАН – филиала ИПМ им. М.В. Келдыша, г. Пущино



Материалы распространяются по лицензии CC-BY

## СОДЕРЖАНИЕ

<b>ВВЕДЕНИЕ</b>	4
<b>Раздел 1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ</b>	6
1.1. Что такое публикация данных	6
1.2 Типы наборов данных в GBIF	7
1.3 Типы лицензий в GBIF	9
1.4 Структурирование и представление метаданных	10
1.5 Стандарт DARWIN CORE для хранения и обмена данными	11
1.6 Публикация данных в Глобальной сети GBIF с использованием IPT	12
<b>Раздел 2. ПРЕДСТАВЛЕНИЕ ДАННЫХ ЛЕТОПИСИ ПРИРОДЫ В DARWIN CORE</b>	13
2.1 Летопись природы и DARWIN CORE	13
2.2 Таксономические списки	14
2.3 Данные научных биологических коллекций	18
2.4 Фенологические наблюдения	23
2.5 Описания, учеты и сборы, выполненные на пробных площадях, маршрутах или трансектах	24
<b>Раздел 3 ВЕРИФИКАЦИЯ, ИСПРАВЛЕНИЕ И УТОЧНЕНИЕ ДАННЫХ</b>	29
3.1 Основные понятия	29
3.2. Технические ошибки	30
3.3. Ошибки согласованности	30
3.4. Исправление ошибок после публикации данных в GBIF.org	31
<b>Раздел 4. МАТЕРИАЛЫ ДЛЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ НА КУРСЕ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ</b>	33
Список литературы	38

## ВВЕДЕНИЕ

Общеизвестно, что “Летопись природы” для заповедников – это основной документ научной деятельности, в котором описываются основные природные процессы и явления, а также изменения в экосистемах заповедника и на сопредельных территориях, вызванные природными и антропогенными факторами в отчетном году. Как правило, в планах научно-исследовательских работ тема “Изучение естественного хода природных процессов и явлений по программе “Летопись природы” идет под номером один, это основная научная работа в заповедниках. Система Российских заповедников уникальна в мире лишь потому, что в заповедниках ведутся ежегодные наблюдения по единой программе.

Собирают материал для Летописи всем миром – научные сотрудники, лаборанты, лесники-наблюдатели, егеря самого заповедника, сотрудники других организаций, аспиранты, студенты, все, кто хотя бы временно работает на его территории. Долгое время считалось, что ведение Летописи природы в системе ООПТ является неотъемлемой частью работы заповедников. С 2007 года, когда был утвержден Приказ Росприроднадзора № 169 от 18.06.2007, Летопись природы начали вести и в российских национальных парках, и сегодня уже накоплен большой объем фактической информации о состоянии природных комплексов.

Ежегодно, после выполнения закрепленных разделов, обработки полевых материалов и утверждения отчетов верстается очередная книга Летописи природы. В ней кратко и понятно приводятся фактические материалы, в основном, в виде таблиц, дополненных краткими комментариями, картосхемами и графиками. Каждая книга Летописи природы составляется в трех идентичных экземплярах. Первый экземпляр хранится в заповеднике как особо ценный документ, второй направляется в вышестоящую организацию, а третий находится в библиотеке заповедника и является рабочим.

Очевидно, что материалы Летописи являются важным источником информации для оценки изменений биоразнообразия. Во многих заповедниках накоплены многолетние (несколько десятилетий) ряды наблюдений, позволяющие анализировать естественную динамику экосистем, выведенных из хозяйственного использования. В то же время, Летопись почти не находит применения (не считая диссертационных, дипломных и курсовых работ) ни на уровне регионов, ни тем более в государственном масштабе. Причина этого кроется в том, что в большинстве заповедников Летопись природы до сих пор хранится только в бумажном виде. Многие уникальные многолетние ряды данных ранее не публиковались в научной литературе и практически недоступны исследователям. Для того чтобы сделать материалы Летописей доступными для научного сообщества, необходима их оцифровка, т.е. представление данных в виде унифицированных электронных таблиц, обнаружимых через Интернет. Это не только откроет новые возможности совместного анализа накопленных данных, но и будет способствовать лучшей сохранности самих сведений, на сбор которых затрачены усилия многих исследователей.

Для отработки механизма оцифровки и публикации данных Летописи природы по рекомендации Минприроды России ФГБУ “Приокско-Террасный государственный заповедник” стал первым российским учреждением, осуществляющим управление особо охраняемыми природными территориями, опубликовавшим данные Летописи природы в глобальной сети открытых данных о биоразнообразии GBIF. Полученный опыт был распространен на другие заповедники и национальные парки благодаря проекту, выполняемому Приокско-Террасным заповедником при грантовой поддержке Секретариата Глобальной информационной системы по биоразнообразию GBIF (проект Russia2019\_03).

В рамках этого проекта был разработан учебный курс и настоящее руководство для обучения сотрудников заповедников и национальных парков оцифровке и публикации данных в сети GBIF. Курс был апробирован с участием представителей, научных сотрудников и заместителей директоров по науке из 18 российских заповедников и национальных парков. В ходе подготовки к проведению курса обучающимися с помощью экспертов были подготовлены и опубликованы более 30 000 записей о находках видов и фенологических явлениях в заповедниках и национальных парках России.

Мы подготовили этот курс для дальнейшего использования, самостоятельного освоения методики и организации подобной школы для сотрудников заповедников и национальных парков.

## Раздел 1. ОСНОВНЫЕ ПОНЯТИЯ И ТЕРМИНЫ

### 1.1. Что такое публикация данных

Все исследователи хорошо знают, что такое публикация научной статьи – текста, в котором рассказывается о методах и результатах проверки тех или иных гипотез, выполненной на основе собранных в природе данных. С развитием сетевых технологий и цифровизации полевой биологии научное сообщество постепенно приходит к пониманию того, что данные сами по себе являются научным продуктом и в составе объединенных массивов могут многократно использоваться для решения самых различных экологических задач (Bisby, 2000). Такое отношение к полевым данным предполагает, что они должны быть обнаружимы через Интернет, доступны в полном объеме и иметь однозначное указание на авторов и источники происхождения, а также правила повторного использования (Wilkinson et al., 2016).

Для быстрого поиска данных и их получения в удобном для пользователя виде существуют специальные хранилища – репозитории. Крупнейшим в мире репозиторием открытых данных о биоразнообразии является Глобальная информационная система GBIF (Global Biodiversity Information Facility, [gbif.org](http://gbif.org)), в настоящее время содержащая более миллиарда записей о находках биологических видов со всего мира (Wheeler, 2004). Публикация через GBIF – это процесс преобразования электронной таблицы с исходными полевыми данными в специальный обменный формат – архив Darwin Core (Wieczorek, et al., 2012), который делает ваши данные доступными через глобальный портал. Все данные в GBIF публикуются в виде отдельных наборов данных (dataset) – логически связанных блоков информации. Например, “Распространение борщевика Сосновского на Европейском Севере”, “Находки птиц в окрестностях г. Пущино”, “Биологические коллекции краеведческого музея поселка Красный Бор” и др. Наборы данных могут быть опубликованы только от имени зарегистрированных в GBIF организаций (publisher). Каждый опубликованный набор имеет авторов, DOI (идентификатор цифрового объекта, Digital Object Identifier) и постоянную страницу на глобальном портале (рис. 1).

Авторы самостоятельно определяют правила использования своих данных, могут вносить изменения и дополнения в опубликованные ими данные, и, при необходимости, отзываться набор данных с публикации. Поскольку GBIF является открытым репозиторием, получить опубликованные данные может любой пользователь. Система GBIF позволяет отслеживать все загрузки опубликованных данных и их цитирование по DOI в научной литературе.

Для обеспечения возможности поиска, объединения и совместного анализа сведений, полученных из GBIF, авторам необходимо предварительно привести свои данные к “общему знаменателю”, единому стандарту оформления данных. В этом пособии приводятся рекомендации по стандартизации данных в соответствии с правилами, принятыми в GBIF. Речь пойдет о материалах Летописи природы и других

сведениях, хранящихся в научных архивах ООПТ. По нашему мнению, даже если вы не планируете в ближайшее время публиковать свои данные через глобальный портал, их оцифровка с использованием международных стандартов позволит вам в будущем существенно сэкономить время при работе с ними.



Рисунок 1. Набор данных о коллекции паукообразных, хранящейся в Приокско-Тerrasном государственном природном биосферном заповеднике (DOI: [10.15468/3cby7](https://doi.org/10.15468/3cby7)). Кнопка CITATIONS открывает список публикаций, в которых использована хотя бы одна запись из этого набора данных.

## 1.2 Типы наборов данных в GBIF

Данные в сети GBIF могут быть опубликованы с разной степенью подробности. В зависимости от того, какие сведения предоставляются и как они логически связаны внутри набора данных, существует четыре типа наборов данных.

**1. Метаданные (Meta Data Only)**, или “данные о данных”, предполагают только описание массива данных без предоставления исходных сведений. Такая публикация позволяет объявить о существовании того или иного массива данных и сделать его потенциально доступным для научного сообщества. Например, можно представить описание гербарной коллекции ООПТ, не имеющей электронного каталога, описать узкоспециализированный тематический ресурс, к которому сложно применить стандарты GBIF, или ресурс, данные которого раскрывать неправомерно, но информация о нем является ценной для специалистов. Последнее касается сводок о распространении редких и охраняемых видов, а также видов, представители которых имеют коммерческую ценность и могут легко стать объектом браконьерства.

Метаданные – это обязательный раздел для всех типов наборов данных, но в данном случае ими и ограничиваются, для остальных трех вариантов предполагается список видов или сведения о находках видов. Схематично типы наборов данных представлены на рис. 2.

## Типы наборов данных, доступных через GBIF

### Метаданные (METADATA only)

+

#### Таксономический список CHECKLIST

для публикации  
таксономических данных:  
списки охраняемых  
видов, тематические  
видовые списки и др.

1 строка – 1 таксон



+

#### Находки OCCURRENCE DATA

Находка - простое полевое  
наблюдение или  
коллекционный образец

1 строка = 1 особь или 1 группа  
особей



+

#### Описания, учеты или сборы SAMPLE EVENT DATA

2 таблицы: данные о сборе  
+ данные о находках

1 запись на листе event =  
1 событие (площадка,  
маршрут)

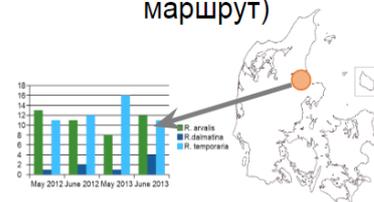


Рисунок 2. Типы наборов данных, доступных через GBIF.org.

**2. Таксономический список (Checklist Data)** содержит список видов и, в некоторых случаях, таксонов более высокого уровня. Таксономический список может быть приведен для какой-либо определенной территории, например, для данного ООПТ, или включать все известные виды для какой-то определенной систематической группы живых организмов. Кроме видовых названий в наборе данных типа Checklist могут быть указаны таксоны более высокого уровня, начиная с того, которому посвящен набор данных.

Если образцы не удалось определить до вида, в таксономический список могут быть включены рода или таксоны высшего ранга. Также надо иметь в виду, что в таксономическом списке определенной группы записи в наборе данных могут начинаться с наименования этой группы, например, отряда, и включать все нижестоящие таксоны, что в данном случае не будет говорить о наличии на данной территории каких-то неопределенных организмов, относящихся к данному таксону. Также можно привести дополнительную информацию, например, названия видов на национальных языках. Самый большой таксономический список в сети GBIF – это GBIF Backbone Taxonomy (DOI: [10.15468/39omei](https://doi.org/10.15468/39omei)), включающий в себя более 6 млн. записей и являющийся справочником, по которому проверяются названия видов во всех наборах данных, публикуемых в GBIF.

**3. Данные о находках (Occurrence Data)** содержат информацию о нахождении того или иного вида (или представителя таксона более высокого уровня) в определенном месте в определенное время, т.е. предполагают наличие сведений о дате находки и географической привязке места нахождения организма (группы)

определенного вида (или отбора образца). В качестве *Occurrence Data* могут быть представлены непосредственно данные о встречах в природе (включая отдельные наблюдения, полученные с помощью автоматических фото- и видеорегистраторов), материалы гербариев, зоологических и других научных коллекций ООПТ, данные фенологических наблюдений, а также опубликованные в научной литературе данные о находках видов, сведения из локальных баз данных, электронных тематических ресурсов и т.д.

**4. Данные описаний, учетов и сборов на пробных площадях, маршрутах и трансектах (*Sampling Event Data*).** В таких наборах данных приводятся сведения о характеристиках сообществ, методах проведения исследований, а также обилии видов, обнаруженных на учетных площадях (маршрутах, трансектах). С помощью *Sampling Event Data* можно представить информацию об урожайности ягодников и продуктивности растительности, учетах птиц на площадках и маршрутах, первичные данные зимних маршрутных учетов, геоботанические описания и т.д. Особая ценность подобных материалов заключается в возможности сохранения первичных данных каждого учета (а не обобщенных суммарных и средних параметров учетов) и последующего сопоставления результатов, полученных одинаковыми методами на разных территориях и / или в разное время.

### 1.3 Типы лицензий в GBIF

Лицензирование в GBIF – это определение правил и требований, согласно которым любые пользователи портала смогут использовать опубликованную вами информацию. При публикации наборов данных авторы должны выбрать одну из общепринятых лицензий. На сегодняшний день система GBIF поддерживает три открытые лицензии, разработанные сообществом Creative Commons. Использование всех, опубликованных через глобальный портал данных и метаданных должно осуществляться по одной из этих трех лицензий.

- CC0 – выбирая эту лицензию, вы делаете ваш набор данных общественным достоянием, пользователи могут использовать информацию по своему усмотрению, при этом ссылка на авторство не обязательна. Рационально использовать эту лицензию, когда нужно максимально широко распространить сведения, например, для публикации таксономических списков, являющихся полностью открытой информацией, собираемой многими специалистами, учеными, инспекторами и др.
- CC-BY – лицензия, описывающая привычные нам правила цитирования научной литературы. Выбирая эту лицензию, вы говорите, что при использовании опубликованных вами метаданных и данных пользователи должны ссылаться на их авторов. Это рекомендуемая нами лицензия для публикуемых находок и коллекций.

- CC-BY-NC – лицензия предусматривает соблюдение тех же правил, что и CC-BY, но вместе с тем ограничивает возможности использования опубликованных метаданных и данных только некоммерческой областью.

#### **1.4 Структурирование и представление метаданных**

Даже если вы не планируете публиковать через глобальный портал описания всех имеющихся массивов данных о биоразнообразии вашей ООПТ, эту информацию полезно структурировать в соответствии с принятыми в GBIF правилами. Это позволит оценить объем ваших данных, представленность информации для разных таксонов или групп организмов, а также определить приоритеты оцифровки данных.

Метаданные каждого набора данных должны включать название и краткое описание (аннотацию) на английском языке. Помимо англоязычной аннотации можно также привести идентичное по содержанию описание на русском языке. Обязательно указать название организации, в которой хранятся данные, контакты авторов, собравших данные, составителей метаданных и создателей набора метаданных.

Помимо описанных выше минимально необходимых метаданных в описание могут быть включены:

- Информация о географическом охвате собранных данных. Она может быть представлена в виде текста или географических координат точек, ограничивающих район исследования.
- Перечень таксонов (любого ранга), входящих в набор исходных неоцифрованных данных.
- Временной интервал сбора данных.
- Ключевые слова.
- Данные о персоналиях, имеющих отношение к сбору данных, но не включенных в список авторов. Например, это может быть список всех коллекторов гербария ООПТ.
- Сведения о проекте (гранте), в процессе выполнения которого были собраны данные.
- Методы сбора данных. Здесь желательно не только описать методы полевых исследований, но и способы фиксации собранного материала, привести ссылки на ключи, использованные для определения образцов, а также описать методы проверки и верификации собранных данных (см. раздел 3).
- Список работ, опубликованных на основе собранных данных.
- Данные о коллекции – название, акроним, сайт, методы хранения образцов и т.д.
- Онлайн ресурсы. В первую очередь – это домашняя страница ресурса, из которого происходит данных набор данных, например, тематическая база данных или информационная система. В ней может содержаться информация, не вошедшая в набор данных, опубликованный через GBIF. Кроме того, в этом

разделе могут быть указаны любые ссылки на ресурсы, каким-либо образом связанные с набором данных.

- Дополнительная информация о наборе данных, такая как логотип, цель создания, периодичность обслуживания (обновления) и др.

Структурированные таким образом метаданные могут храниться как в привычном текстовом формате, так и в электронной таблице или базе данных. При необходимости они могут быть легко опубликованы через GBIF.org.

При публикации наборов исходных данных (см. ниже) метаданные оформляются по таким же правилам.

### 1.5 Стандарт DARWIN CORE для хранения и обмена данными

Для объединения первичных данных о биоразнообразии из разных источников требуется единообразное оформление или, иными словами, приведение их к общему стандарту. Стандарт – это общепринятый набор правил, многими из которых мы пользуемся ежедневно. Алфавит, цифры, почтовые индексы, автомобильные номера – все это примеры стандартов из повседневной жизни. При работе с оцифрованными данными о биологическом разнообразии под стандартом понимается набор полей (столбцов) в таблице и правила их заполнения.

Наиболее распространенным в GBIF в настоящее время является стандарт Darwin Core (сокращенно DwC), разрабатываемый рабочей группой [Biodiversity Informatics Standards](#), ранее известной как Taxonomic Databases Working Group (TDWG). Этот стандарт включает в себя более 200 терминов, т.е. возможных полей (заголовков столбцов) для хранения цифровых данных о биоразнообразии. Полный перечень терминов и их спецификация (подробное описание) на английском языке доступны на странице [Darwin Core quick reference guide](#). В следующих разделах данного пособия приводится описание и правила использования наиболее употребительных терминов.

Для удобства пользователей на сайте термины DwC сгруппированы в несколько тематических разделов. Раздел **Record-level Terms** для описания записи: тип, название набора данных, код коллекции, информация об организации, в которой хранятся данные (образцы) и прочее. В разделе **Occurrence** приводятся идентификатор образца в коллекции или код записи полевых наблюдений, а также сведения о состоянии организма во время находки или сбора (численность, пол, репродуктивное состояние, поведение, жизненная стадия и др.). Кроме того, в этом разделе можно указать ссылки на фотографию. Время и методы сборов описываются терминами раздела **Event**, здесь также может быть представлена информация о биотопе. В разделе **Location** можно подробно охарактеризовать географическую привязку места находки. Таксономическое положение вида (образца) описывается в разделе **Taxon**, информация о специалисте, сделавшем определение, а также источники, использованные для этого – в разделе **Identification**. Описание особенностей конкретного организма или таксономически однородной группы

организмов приводится в разделе **Organism**. Любая дополнительная информация может быть представлена в терминах "...Remarks" соответствующих разделов.

Концепция DwC предполагает хранение как исходной, так и формализованной информации с указанием методов формализации. Для хранения исходной информации используются термины группы "*verbatim*", в буквальном переводе "дословно", (например, *verbatimEventDate*, *verbatimLocality* и др.); информация в них должна в точности соответствовать первоисточнику (гербарной этикетке, полевому дневнику, литературной публикации и т.п.). Термины, предназначенные для хранения формализованной информации, часто предполагают выбор значения из фиксированного списка или использование определенных правил для представления данных. Например, для термина *disposition*, описывающего актуальное состояние коллекционного образца, возможны значения "*in collection*", "*missing*", "*voucher elsewhere*", "*duplicates elsewhere*". Для даты и времени используется международный стандарт ISO 8601:2004(E). Данные в полях "*verbatim*" следует приводить на языке оригинала, а формализованную информацию – желательно на английском языке.

Текущая версия DwC содержит более 200 терминов, позволяющих описать данные, полученные из разных источников. Более 30 расширений (*extensions*) DwC для специфических данных позволяют описывать характеристики отобранных проб, результаты молекулярно-генетического анализа, аудио- и видеоматериалы, и многие другие атрибуты, обычно не применимые к большинству данных, но в определенных областях исследований имеющие большую ценность. С помощью терминов DwC также можно описывать данные о локалитетах, где целевые виды не были обнаружены, что важно для решения задач мониторинга популяций охраняемых и инвазионных видов и изучения динамики их ареалов.

## 1.6 Публикация данных в Глобальной сети GBIF с использованием IPT

Процедура публикации данных через GBIF.org состоит из следующих этапов: (1) регистрация организации в GBIF, (2) выбор типа набора данных, (3) составление описания набора данных – метаданных, (4) стандартизация исходных данных, если публикация не ограничивается метаданными, и (5) собственно публикация набора данных как технический процесс.

Непосредственно публикация данных в GBIF – это техническая процедура, в ходе которой метаданные и электронная таблица со стандартизированными согласно требованиям Darwin Core исходными данными преобразуются в специальный обменный формат – архив Darwin Core, затем набор данных регистрируется и индексируется в системе GBIF и становится общедоступным через глобальный портал. Для создания архива Darwin Core существуют различные программные решения, наиболее популярным является IPT – Integrated Publishing Toolkit (Robertson et al., 2014). Для того, чтобы опубликовать набор данных, нужно установить собственный IPT или прикрепить свою организацию к любой из уже существующих инсталляций, связавшись с администратором выбранной IPT. В любом случае данные будут опубликованы с сохранением авторства и от имени организации, из которой они

происходят. Учётные записи в IPT могут иметь сразу несколько сотрудников одной организации. Все опубликованные данные хранятся на сервере, где установлена IPT, к которой “привязана” организация, портал GBIF.org не хранит исходные данные, а только индексирует их, поэтому ответственность за резервное копирование опубликованных наборов данных лежит на администраторе IPT (Шашков и др., 2017). В ходе реализации проекта для публикации материалов Летописей природы использована IPT, установленная в Институте математических проблем биологии – филиале ИПМ им. М.В. Келдыша РАН (Шашков, 2018). Вы также можете использовать эту инсталляцию, связавшись с ее администраторами по адресу [gbif.ru@yandex.ru](mailto:gbif.ru@yandex.ru).

Чтобы опубликовать набор данных через IPT, необходимо войти в свой аккаунт в IPT и создать новый набор данных соответствующего типа. Затем следует загрузить файл с исходными данными, соотнести заголовки полей исходной таблицы с терминами Darwin Core и внести метаданные. После этого вы можете сгенерировать (автоматически) архив Darwin Core и зарегистрировать свой набор данных в сети GBIF. Подробная инструкция по публикации набора данных через IPT доступна архиве с учебными материалами.

## Раздел 2. ПРЕДСТАВЛЕНИЕ ДАННЫХ ЛЕТОПИСИ ПРИРОДЫ В DARWIN CORE

### 2.1 Летопись природы и DARWIN CORE

В этом разделе пособия представлены методические рекомендации по оцифровке разных типов данных, содержащихся в Летописи природы в соответствии со стандартом Darwin Core. В каждом подразделе перечислены обязательные и рекомендуемые для заполнения поля. Для каждого термина стандарта Dwc приводится название, описание его использования и ссылка на оригинальную спецификацию (на английском языке) на сайте разработчиков стандарта – рабочей группы TDWG <https://dwc.tdwg.org/terms/>. Также даны практические советы по форматированию данных.

Летопись природы – это научная программа, преследующая цель систематического сбора данных по динамике природных явлений и процессов в пределах заповедника в течение неопределенно длительного времени. Согласно методическим указаниям (Филонов, Нухимовская, 1990) эта программа включает следующие этапы:

**1. Инвентаризация видового состава растений, грибов и животных ООПТ.** В Dwc эти данные можно оформить как таксономические списки (checklist data);

**2. Выполнение систематических наблюдений за отдельными объектами природы.** Согласно структуре, предложенной Филоновым, Нухимовской (1990), под эту категорию попадают данные из следующих разделов Летописи природы:

- новые виды и встречаемость редких и исчезающих видов животных и растений на всей территории заповедника, места встреч птичьих гнезд и др. (данные из карточек встреч);

- результаты учетов с картографированием на разовых и постоянных маршрутах, например, куриных и водоплавающих птиц и других животных, по которым имеются данные о месте обнаружения вида;
- фенологический календарь природы на базе регулярных фенологических наблюдений за растениями, грибами и животными, в том числе на площадках и маршрутах;
- создание и пополнение гербариев, зоологических и других коллекций.

В соответствии со стандартом DwC вся эта информация может быть представлена как данные о находках (Occurrence Data).

**3. Картографирование растительности и животного населения на пробных площадях, изучение сообществ и определение параметров сообществ.** Под эту категорию попадают данные, собранные на постоянных пробных площадях (ППП) и маршрутах из следующих разделов Летописи природы:

- растительность и её изменения (состав и структура сообществ, продуктивность и урожайность);
- учеты численности позвоночных животных на ППП и постоянных маршрутах, линиях давилок и канавок, численность рыб и беспозвоночных животных численность и т.п.

В соответствии со стандартом DwC это “описания, учеты и сборы” (Sampling Event Data). Особая ценность подобных материалов заключается в возможности сохранения данных каждого учета (а не обобщенных суммарных и средних параметров), что дает возможность последующего сопоставления результатов, полученных одинаковыми методами на разных территориях и / или в разное время. К сожалению, по некоторым разделам в Летописи природы не предусматривается предоставление первичной информации, так как форматы таблиц предполагают частично обработанную информацию. Например, по результатам зимних маршрутных учетов и учетов птиц на маршрутах в Летопись включались уже обобщенные данные, а первичные данные каждого учета хранятся в архивах и полевых дневниках исследователей. Если эти данные не удалось сохранить, то результаты зимних маршрутных учетов и аналогичных иных учетных работ не могут быть представлены в стандарте DwC.

## 2.2 Таксономические списки

Данные инвентаризации биологического разнообразия ООПТ следует представлять в виде Checklist Data. Для оцифровки этих данных необходимо использовать поля из раздела Taxon стандарта Darwin Core (<https://dwc.tdwg.org/terms/#taxon>), а при их публикации через IPT выбирать ядро (Core) Checklist. Одна запись (строка) в таблице с данными должна соответствовать одному таксону, таксоны не должны повторяться внутри одной таблицы. Таблица с таксономическими списками должна включать три обязательных поля (столбца), представленных в табл. 1.

Таблица 1. Обязательные поля для представления данных таксономических списков.

taxonID	Уникальный <i>идентификатор таксона</i> , который в идеале должен быть глобальным и связанным не только с информацией о данном таксоне, но также и с дополнительной информацией, такой как изображения, видео и пр. Удобно использовать идентификаторы таксонов, используемых самой системой GBIF. О получении этих идентификаторов написано ниже. <a href="https://dwc.tdwg.org/terms/#taxonID">https://dwc.tdwg.org/terms/#taxonID</a>
scientificName	Полное, насколько это возможно, <i>научное название таксона</i> , в случае если это вид включающее род, видовой эпитет, авторский знак и год его описания. Корректными являются названия, соответствующие правилам для соответствующей группы таксонов. Нельзя приводить рабочие названия (“ <i>Asplenium sp2</i> ”), названия с пометками о точности определения (“ <i>Anemone cf. nemorosa</i> ”), или названия на русском языке (“рыбный филин”). Для последних предусмотрено специальное поле vernacularName ( <a href="https://dwc.tdwg.org/terms/#vernacularName">https://dwc.tdwg.org/terms/#vernacularName</a> ) <a href="https://dwc.tdwg.org/terms/#dwc:scientificName">https://dwc.tdwg.org/terms/#dwc:scientificName</a>
taxonRank	<i>Ранг таксона</i> , до которого удалось определить образец (вид, род, семейство и т.д.). Это поле необходимо для корректной индексации набора данных на глобальном портале и сопоставления ваших названий с таксономией GBIF. Для заполнения этого поля необходимо использовать линнеевские ранги: царство, тип, класс, порядок, семейство, род, вид. Допускается написание рангов на латинском или английском языке. <a href="https://dwc.tdwg.org/terms/#dwc:taxonRank">https://dwc.tdwg.org/terms/#dwc:taxonRank</a>

Помимо обязательных для заполнения полей настоятельно рекомендуется также заполнить следующие (табл. 2):

Таблица 2. Рекомендуемые поля для представления данных таксономических списков.

kingdom	<i>Царство</i> необходимо указывать для однозначного сопоставления ваших названий с таксономией GBIF. Как известно, существуют виды с одинаковыми названиями и относящиеся к разным таксономическим группам. Кроме того, в названиях ваших таксонов возможны опечатки, или другие неточности. Для заполнения этого поля необходимо указывать линнеевские ранги (“ <i>Animalia</i> ”), нельзя указывать нетаксономические группы (“ <i>Algae</i> ” или “ <i>Herbivora</i> ”) <a href="https://dwc.tdwg.org/terms/#dwc:kingdom">https://dwc.tdwg.org/terms/#dwc:kingdom</a>
parentNameUsageID	<i>Идентификатор таксона следующего (более высокого) ранга</i> . Необходим для поддержания корректной таксономической иерархии в ваших данных. <a href="https://dwc.tdwg.org/terms/#dwc:parentNameUsageID">https://dwc.tdwg.org/terms/#dwc:parentNameUsageID</a>
acceptedNameUsageID	<i>Идентификатор валидного названия вида</i> . Его необходимо указывать, если вы используете в наборе данных синонимичное название, а не общепринятое. <a href="https://dwc.tdwg.org/terms/#dwc:acceptedNameUsageID">https://dwc.tdwg.org/terms/#dwc:acceptedNameUsageID</a>

Настоятельно рекомендуем перед подготовкой таблицы проверить ваши названия на соответствие таксономической системе GBIF. Это также позволит

получить идентификаторы таксонов и проверить ваши данные на предмет орфографических ошибок. Существует множество веб-инструментов для проверки таксономических названий, мы предлагаем воспользоваться Species matching, доступным через портал GBIF.org в разделе Tools (рис. 3).

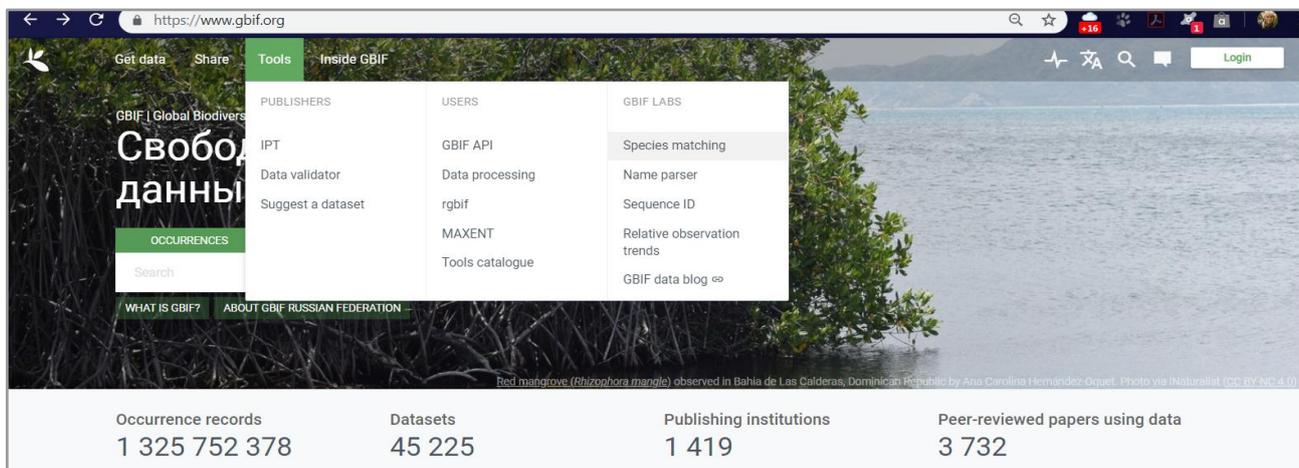


Рисунок 3. Инструмент Species matching на портале GBIF.org.

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).

В открывшемся окне вы увидите список ваших таксонов. Если возможно, укажите Царство, к которому они относятся (кликнув соответствующую картинку), после чего нажмите кнопку MATCH TO GBIF BACKBONE (рис. 4).

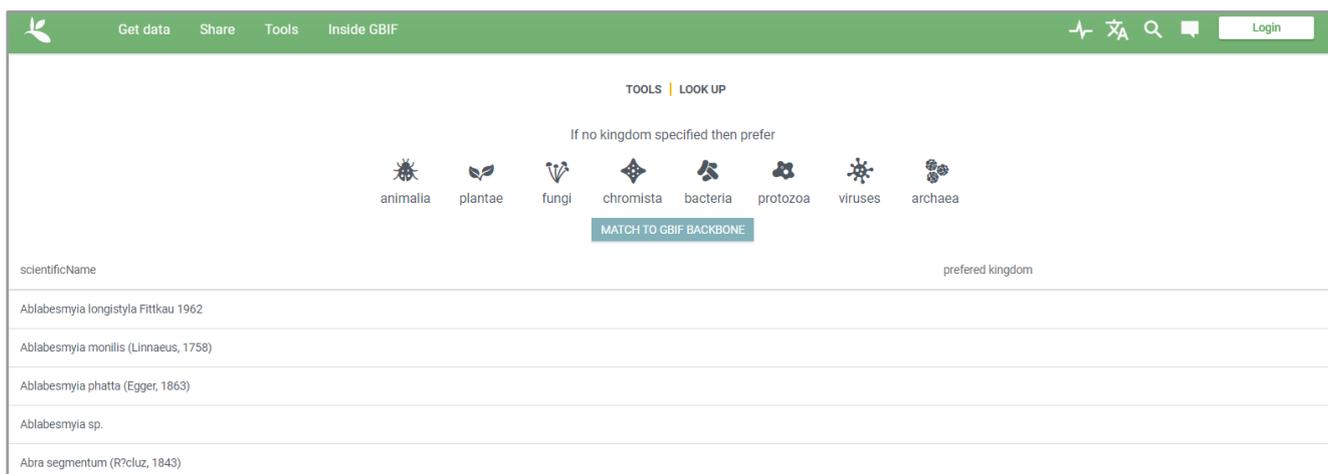


Рисунок 4. Интерфейс инструмента Species matching с загруженным для проверки таксономическим списком.

Результат сопоставления доступен для просмотра в браузере (рис. 5), также его можно сохранить в CSV файл, который будет содержать следующие поля:

**occurrenceId** – идентификатор. Если вы его не указывали, поле будет пустым;

**verbatimScientificName** – указанные 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>. Эту информацию (значение key или URL) вы можете использовать в качестве идентификатора taxonID;

**matchType** – результат сопоставления. В этом поле возможны три значения: EXACT – название полностью соответствует GBIF Backbone, FUZZY – т.н. подозрительные, название не полностью соответствует GBIF Backbone, есть расхождения в нескольких буквах, возможно вследствие опечатки в исходных данных, HIGHERRANK – таксон не найден в GBIF Backbone, но есть родительский таксон более высокого ранга, например, вид отсутствует в GBIF, но есть соответствующий род;

**confidence** – количественная характеристика степени совпадения в процентах;

**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.

В следующих столбцах представлено положение таксона в иерархическом порядке, начиная с царства.

verbatimScientificName	preferredKingdom	matchType	confidence	scientificName (editable)	status	rank	kingdom
Ablabesmyia longistyla Fittkau 1962	animalia	EXACT	100	Ablabesmyia longistyla Fittkau, 1962	ACCEPTED	Species	<a href="#">Animalia</a>
Ablabesmyia monilis (Linnaeus, 1758)	animalia	EXACT	100	Ablabesmyia monilis (Linnaeus, 1758)	ACCEPTED	Species	<a href="#">Animalia</a>
Ablabesmyia phatta (Egger, 1863)	animalia	EXACT	100	Ablabesmyia phatta (Egger, 1863)	ACCEPTED	Species	<a href="#">Animalia</a>
Ablabesmyia sp.	animalia	EXACT	85	Ablabesmyia Johannsen, 1905	ACCEPTED	Genus	<a href="#">Animalia</a>
Abra segmentum (Récluz, 1843)	animalia	EXACT	100	Abra segmentum (Récluz, 1843)	ACCEPTED	Species	<a href="#">Animalia</a>
Acarina sp.	animalia	EXACT	84	Acarina Baraud, 1965	SYNONYM	Genus	<a href="#">Animalia</a>
Acilius sp.	animalia	EXACT	80	Acilius Leach, 1817	ACCEPTED	Genus	<a href="#">Animalia</a>
Acricotopus lucens (Zetterstedt, 1850)	animalia	EXACT	100	Acricotopus lucens (Zetterstedt, 1850)	ACCEPTED	Species	<a href="#">Animalia</a>
Acricotopus sp.	animalia	EXACT	85	Acricotopus Kieffer, 1921	ACCEPTED	Genus	<a href="#">Animalia</a>

Рисунок 5. Результат проверки таксономического списка.

Названия таксонов можно отредактировать в полученном файле CSV, или непосредственно в браузере, а затем сохранить результат в хфайл. Для редактирования онлайн необходимо нажать на требующее корректировки название в поле scientificName (со значком в виде карандаша) и в открывшемся диалоговом окне выбрать корректное название (рис. 6) из предложенных автоматически или тех, которые появятся, если вы начнете вводить другой вариант.

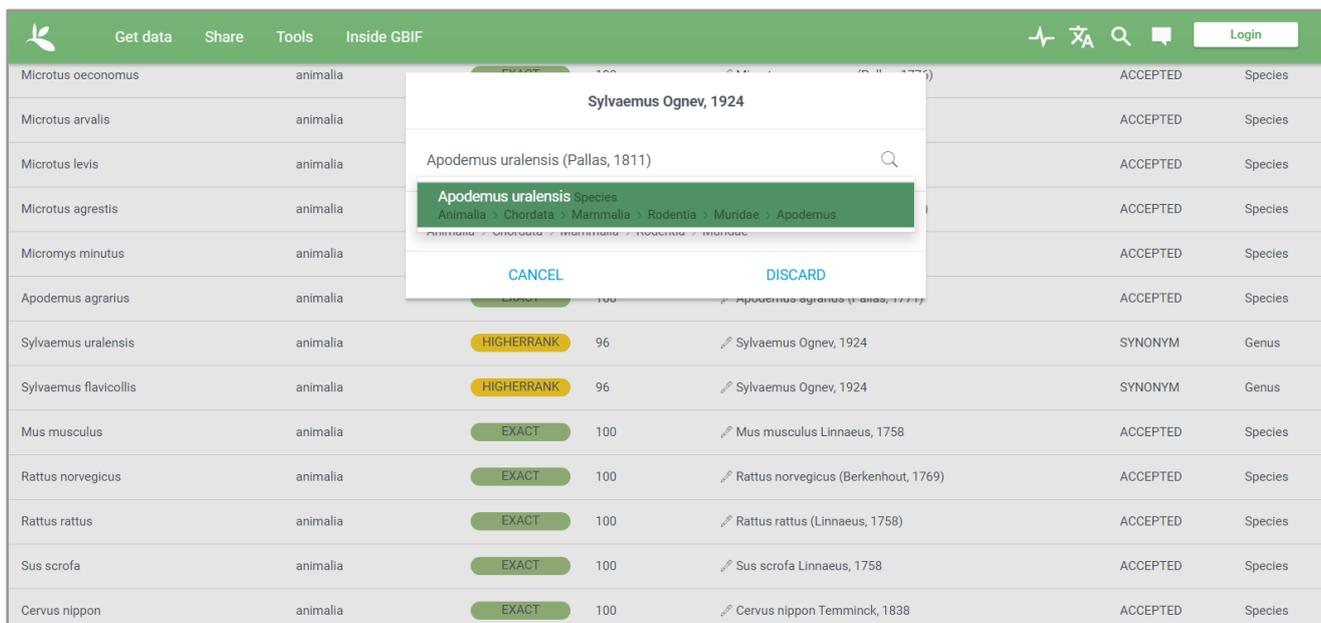


Рисунок 6. Редактирование исходных названий таксонов через веб-интерфейс.

Таким образом, с помощью инструмента Species matching вы можете проверить соответствие названий вашего списка видов таксономической системе GBIF Backbone, выявить наличие орфографических ошибок, при необходимости скорректировать названия и получить данные для заполнения обязательных полей taxonID и taxonRank.

**Совет.** Если в ваших исходных данных сведения о роде и виде хранятся в разных столбцах электронной таблицы Excel, их можно легко объединить, воспользовавшись формулой =A2&" "&B2 (рис. 7).

	A	B	C	D
1	GENUS	SPECIES		
2	Strix	albitarsis	=A2&" "&B2	
3	Strix	capensis	Strix capensis	
4	Strix	uralensis	Strix uralensis	
5	Strix	woodfordii	Strix woodfordii	
6	Strix	nebulosa	Strix nebulosa	
7				

Рисунок 7. Объединение столбцов с использованием формулы EXCEL.

### 2.3 Данные научных биологических коллекций

В этом разделе как оцифровку коллекций мы будем рассматривать перенос информации с этикеток в электронные таблицы с полями DwC, а не получение изображений образцов. Последнее, безусловно, существенно дополняет табличные

данные и расширяет возможности использования коллекции (Шашков и др., 2018), но требует значительных трудовых, аппаратных и финансовых ресурсов.

Данные биологических коллекций могут быть представлены в виде Occurrence Data. Для переноса данных биологических коллекций в электронные таблицы DwC могут использоваться термины из разных разделов стандарта, но базовая информация должна отвечать на вопросы **где, когда и кем** собран образец, а также **что** это такое. Одна запись в таблице должна соответствовать одному образцу в коллекции, собранному в определенном месте и времени. Очевидно, что образцы одних и тех же видов могут быть собраны в разных местах и неоднократно за период исследования. При публикации таких данных через IPT необходимо выбирать ядро (Core) Occurrences.

Для представления данных об образцах биологических коллекций стандартом DwC предусмотрено 4 обязательных для заполнения поля (табл. 3).

Таблица 3. Обязательные поля для представления данных об образцах биологических коллекций.

occurrenceID	<i>Уникальный идентификатор находки.</i> Вы можете составить его самостоятельно, или сгенерировать глобальный идентификатор GUID ( <a href="https://www.guidgenerator.com/">https://www.guidgenerator.com/</a> ). Желательно при составлении ID не ограничиваться сплошной нумерацией записей (1,2,3,4,5 и т.д.), а включить в идентификатор акроним коллекции или аббревиатуру ООПТ, каталожный номер образца или другую подобную информацию. <a href="https://dwc.tdwg.org/terms/#dwc:occurrenceID">https://dwc.tdwg.org/terms/#dwc:occurrenceID</a>
basisOfRecord	В этом поле необходимо указать, <i>что послужило основанием для появления этой записи.</i> Для его заполнения предусмотрен фиксированный набор значений, необходимо использовать одно из них – PreservedSpecimen, FossilSpecimen, LivingSpecimen, MaterialSample, Event, HumanObservation, MachineObservation, Taxon, Occurrence. В случае коллекций необходимо выбирать значение <b>PreservedSpecimen</b> , при этом для всех терминов и значений важно сохранить оригинальное (слитное) написание. <a href="https://dwc.tdwg.org/terms/#dwc:basisOfRecord">https://dwc.tdwg.org/terms/#dwc:basisOfRecord</a>
scientificName	<i>Научное название таксона.</i> Заполняется по тем же правилам, что и для таксономических данных. Настоятельно рекомендуется проверить все названия с помощью инструмента <a href="#">Species matching</a> (подробнее см. раздел 2.2) <a href="https://dwc.tdwg.org/terms/#dwc:scientificName">https://dwc.tdwg.org/terms/#dwc:scientificName</a>
eventDate	<i>Дата сбора образца.</i> Должна быть представлена в формате ГГГГ-ММ-ДД, т.е. дата 27 июня 2019 года в DwC будет иметь вид 2019-06-27. <a href="https://dwc.tdwg.org/terms/#dwc:eventDate">https://dwc.tdwg.org/terms/#dwc:eventDate</a>

**Совет.** Чтобы привести даты к требуемому формату, необходимо в таблице Excel выделить соответствующую ячейку (или весь столбец), кликнув правой кнопкой мыши, выбрать пункт меню Формат ячеек -> Все форматы, задать необходимый формат даты вручную в поле “Тип”, нажать ОК (рис. 8). Окно формата ячеек можно также вызвать сочетанием клавиш Ctrl+1.

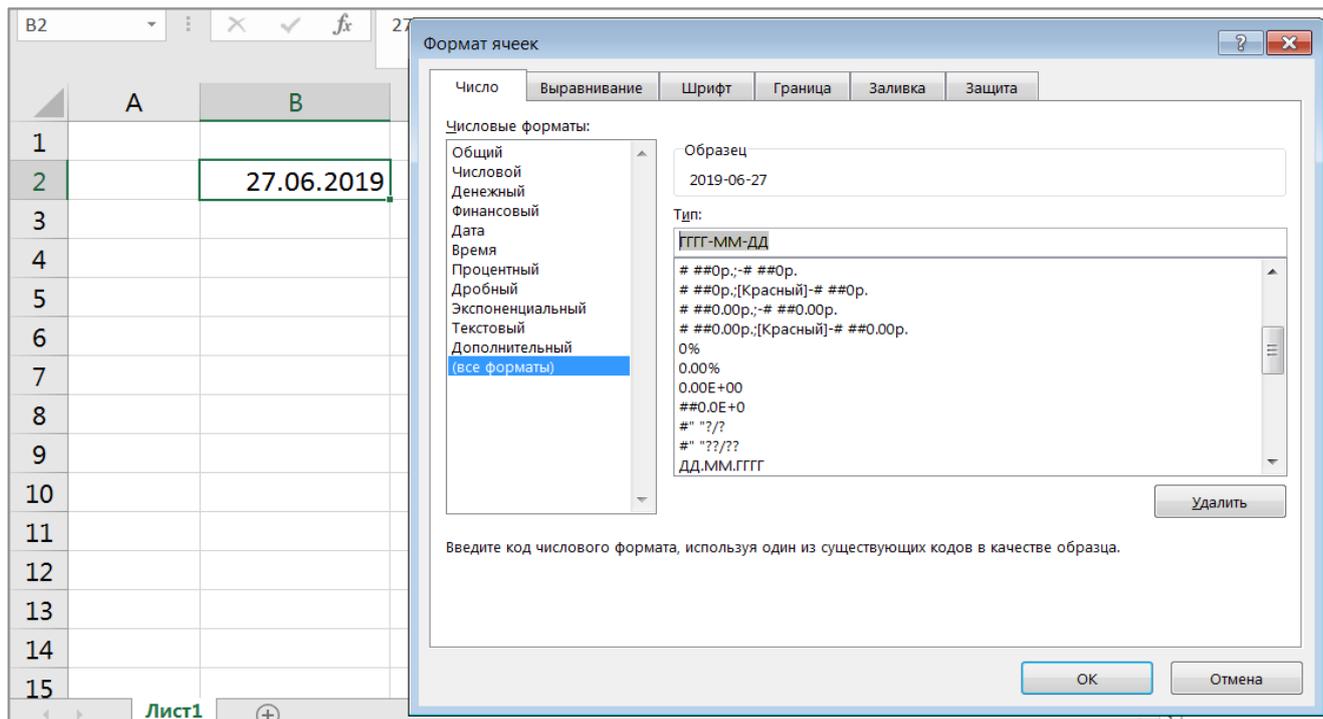


Рисунок 8. Задание формата даты ГГГГ-ММ-ДД в Excel.

Очевидно, что информация об образцах, хранящихся в коллекциях, которую можно прочитать на этикетках, гораздо обширнее четырёх полей, представленных в таблице выше. В табл. 4 приводятся наиболее употребительные термины стандарта DwC для оцифровки данных этикеток научных биологических коллекций.

Таблица 4. Наиболее часто используемые термины DwC для представления этикеточных данных биологических коллекций.

institutionCode	Сокращенное название организации, в которой хранится коллекция <a href="http://rs.tdwg.org/dwc/terms/institutionCode">http://rs.tdwg.org/dwc/terms/institutionCode</a>
collectionCode	Акроним коллекции <a href="http://rs.tdwg.org/dwc/terms/collectionCode">http://rs.tdwg.org/dwc/terms/collectionCode</a>
recordedBy	В этом поле на первом месте указывается коллектор, а далее все лица, имеющие отношение к данной находке. Имена следует отделять вертикальной чертой, например Oleg Borodin   Alexey Petrovich Seregin <a href="http://rs.tdwg.org/dwc/terms/recordedBy">http://rs.tdwg.org/dwc/terms/recordedBy</a>
identifiedBy	В этом поле приводится список лиц, определивших образец. Правила заполнения аналогичны предыдущему полю. <a href="http://rs.tdwg.org/dwc/terms/identifiedBy">http://rs.tdwg.org/dwc/terms/identifiedBy</a>
taxonRank	Ранг таксона, до которого удалось определить образец (вид, род, семейство и др.). Это поле необходимо для корректной индексации набора данных на глобальном портале и сопоставления ваших названий с таксономией GBIF. Для заполнения этого поля необходимо использовать линеевские ранги: царство, тип, класс, порядок, семейство, род, вид. Допускается написание на латинском или английском языке. <a href="https://dwc.tdwg.org/terms/#dwc:taxonRank">https://dwc.tdwg.org/terms/#dwc:taxonRank</a>

kingdom	<i>Царство</i> . Его необходимо указывать для однозначного сопоставления ваших названий с таксономией GBIF. Как известно, существуют виды с одинаковыми названиями и относящиеся к разным таксономическим группам. Кроме того, в названиях ваших таксонов возможны опечатки, или другие неточности. Для заполнения этого поля необходимо указывать линнеевские ранги (“Animalia”), нельзя указывать нетаксономические группы (“Algae” или “Herbivora”) <a href="https://dwc.tdwg.org/terms/#dwc:kingdom">https://dwc.tdwg.org/terms/#dwc:kingdom</a>
countryCode	<i>Код страны</i> согласно международному стандарту ISO 3166-1-alpha-2. Его необходимо приводить для однозначного указания страны. Код России – RU. <a href="http://rs.tdwg.org/dwc/terms/countryCode">http://rs.tdwg.org/dwc/terms/countryCode</a>
stateProvince	<i>Административно-территориальная единица</i> страны. Для России в этом поле указывают название области, края или республики. <a href="http://rs.tdwg.org/dwc/terms/stateProvince">http://rs.tdwg.org/dwc/terms/stateProvince</a>
county	<i>Административно-территориальная единица следующего (более дробного) уровня</i> . Как правило, в этом поле указывают название района. <a href="http://rs.tdwg.org/dwc/terms/county">http://rs.tdwg.org/dwc/terms/county</a>
locality	<i>Место сбора</i> . Обычно указывается расстояние и направление до населенного пункта. Информацию в этом поле лучше приводить на английском языке. Важно сохранить и русскоязычное описание места сбора, такую информацию следует поместить в поле verbatimLocality. <a href="http://rs.tdwg.org/dwc/terms/locality">http://rs.tdwg.org/dwc/terms/locality</a>
verbatimLocality	<i>Описание места сбора в соответствии с исходными данными</i> . Приводится на языке оригинала в точном соответствии этикетке. <a href="http://rs.tdwg.org/dwc/terms/verbatimLocality">http://rs.tdwg.org/dwc/terms/verbatimLocality</a>
decimalLatitude	<i>Географическая широта в десятичных градусах</i> . Для северного полушария значения широты положительные, для южного – отрицательные. <a href="http://rs.tdwg.org/dwc/terms/decimalLatitude">http://rs.tdwg.org/dwc/terms/decimalLatitude</a>
decimalLongitude	<i>Географическая долгота в десятичных градусах</i> . Для восточного полушария значения долготы положительные, для западного – отрицательные. <a href="http://rs.tdwg.org/dwc/terms/decimalLongitude">http://rs.tdwg.org/dwc/terms/decimalLongitude</a>
geodeticDatum	<i>Эллипсоид, геодезическая система координат или система пространственной привязки (SRS)</i> , на которой основаны географические координаты. <a href="http://rs.tdwg.org/dwc/terms/geodeticDatum">http://rs.tdwg.org/dwc/terms/geodeticDatum</a>
coordinatePrecision	<i>Точность определения координат</i> , выраженная в долях градуса. Например, для показаний спутникового навигатора 54.75375, 35.36574 точность составляет 0.00001; точность координат 54, 35 составляет 1.0. Значение 0.01667 соответствует одной угловой минуте, 0.000278 – одной секунде. <a href="http://rs.tdwg.org/dwc/terms/coordinatePrecision">http://rs.tdwg.org/dwc/terms/coordinatePrecision</a>

coordinateUncertaintyInMeters	Оценка погрешности определения координат, выраженная в метрах. Этот термин очень важен, т.к. координаты могут быть определены с высокой точностью (5 знаков после запятой), но реальное место сбора не всегда будет соответствовать этим координатам. Например, при использовании обычного туристического GPS приемника погрешность определения координат, как правило, составляет 3–5 метров. При определении координат по топографическим картам общедоступных масштабов 1:200 000 или 1:500 000, погрешность привязки, в лучшем случае, составляет сотни метров. Если значение погрешности неизвестно, или его невозможно оценить (например, когда координаты не указаны), поле следует оставить пустым. "0" не является допустимым значением. <a href="http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters">http://rs.tdwg.org/dwc/terms/coordinateUncertaintyInMeters</a>
individualCount	Число обнаруженных особей, выраженное в штуках. Например, <b>37</b> . <a href="http://rs.tdwg.org/dwc/terms/individualCount">http://rs.tdwg.org/dwc/terms/individualCount</a>
organismQuantity	Этот термин используется, если для оценки обилия выбрано не число особей, а другая величина. В этом случае значение обилия помещают в этом поле. <a href="http://rs.tdwg.org/dwc/terms/organismQuantity">http://rs.tdwg.org/dwc/terms/organismQuantity</a>
organismQuantityType	Размерность величины, указанной в предыдущем поле. Например, <i>r</i> в organismQuantity и <b>Braun Blanquet Scale</b> в organismQuantityType. <a href="http://rs.tdwg.org/dwc/terms/organismQuantityType">http://rs.tdwg.org/dwc/terms/organismQuantityType</a>
year	Год сбора образца (4 цифры) <a href="http://rs.tdwg.org/dwc/terms/year">http://rs.tdwg.org/dwc/terms/year</a>
month	Месяц сбора образца (1...12) <a href="http://rs.tdwg.org/dwc/terms/month">http://rs.tdwg.org/dwc/terms/month</a>
day	День сбора образца <a href="http://rs.tdwg.org/dwc/terms/day">http://rs.tdwg.org/dwc/terms/day</a>

**Совет.** Если ваши координаты представлены в формате ГГ ММ СС, их можно пересчитать в десятичные градусы, используя формулу:

$$\text{ГГ.ГГГГГ} = \text{ГГ} + \text{ММ}/60 + \text{СС}/3600$$

Пересчет легко выполнить, используя формулы в электронных таблицах (рис. 9), но важно понимать, что в этом случае значения должны храниться в отдельных полях (столбцах), без знаков .°, ' и ".

	A	B	C	D	E	F	G	H	I
1	DD_lat	MM_lat	SS_lat	DD_lon	MM-lon	SS_lon	DD.DDDDD_lat	DD.DDDDD_lon	
2	50	40	45	40	50	30.3	50.07917	40.84175	
3									
4									
5									

Рисунок 9. Пересчет координат в десятичные градусы с использованием формул EXCEL.

На портале Gis-lab доступна готовая таблица с формулами для пересчета координат в десятичные градусы (<http://gis-lab.info/qa/dms2dd.html>). Вам потребуется только ввести координаты в таблицу с готовыми формулами, а затем – скопировать результаты в таблицу с вашими данными.

Если в исходных данных координаты хранятся в ячейках целиком, для конвертации их в десятичные градусы удобнее воспользоваться веб-инструментом национального портала о биоразнообразии Канады Canadensys <http://data.canadensys.net/tools/coordinates?lang=fr>. Для расчета используется та же формула, что и в предыдущем примере, но входные данные могут содержать знаки °, ' и ''.

Вычисленные десятичные координаты следует округлить, чтобы они соответствовали точности определения координат (coordinatePrecision). Избыточное число знаков после запятой не несёт никакой полезной информации, а только вводит в заблуждение.

## 2.4 Фенологические наблюдения

Данные фенологических наблюдений можно представить также в виде Occurrence Data. Как и в случае с данными о коллекционных образцах, таблица должна обязательно включать поля occurrenceID, basisOfRecord, scientificName и eventDate (см. табл. 3).

Отметим, что стандартом DwC пока не предусмотрено специальных полей, предназначенных для фенологических данных. Частично эту информацию можно описать с помощью терминов lifeStage (<http://rs.tdwg.org/dwc/terms/#dwciri:lifeStage>), в который включают информацию о классе возраста (онтогенетическом состоянии) наблюдаемого объекта, а также behavior (<http://rs.tdwg.org/dwc/terms/#dwc:behavior>), позволяющего описывать поведение объекта во время находки (встречи). Тем не менее, эти два термина не позволяют описать все разнообразие наблюдаемых по программе Летописи природы фенологических явлений, кроме того, разнесение однородных по смыслу данных в разные столбцы противоречит логике представления этой информации. В результате обсуждений с участниками проекта ECN и службой технической поддержки GBIF сформулированы следующие варианты представления данных о фенологических явлениях с помощью стандарта DwC.

**Вариант 1.** Использовать термин dynamicProperties (<http://rs.tdwg.org/dwc/terms/dynamicProperties>). В это поле вносят данные о всех измерениях и характеристиках объекта, которые были описаны во время его наблюдения. Для представления информации необходимо использовать схему кодирования “ключ : значение” в соответствии с нотацией JSON. Оформленные таким способом данные будут “машиночитаемыми”, т.е. напоминать фрагмент программного кода. Например, явление “первая песня” в таком формате будет иметь вид {"PhenologicalEventType": "First song"}. Вероятно, это не самый удобный вариант для хранения таких данных в электронных архивах ООПТ.

**Вариант 2.** Использование раздела Dwc MeasurementOrFact <http://rs.tdwg.org/dwc/terms/#measurementorfact>. Это раздел позволяет подробно описать качественные и количественные характеристики объекта во время наблюдений в “человекочитаемом” формате. Для фенологических данных могут быть использованы термины, перечисленные в табл. 5. В этом случае мы рассматриваем фенологическое явление как качественную переменную, а сами явления – это значения, которые она может принимать.

Таблица 5. Термины раздела MeasurementOrFact, используемые для представления данных Летописи природы.

measurementType	<i>Измеряемая (оцениваемая) величина.</i> Для данных о фенологических явлениях в этом поле везде нужно указать “phenological event type” <a href="http://rs.tdwg.org/dwc/terms/measurementType">http://rs.tdwg.org/dwc/terms/measurementType</a>
measurementValue	<i>Значение измеряемой величины.</i> В случае фенологических явлений этим значением будет тип фенологического явления “начало зацветания”, “первая песня” и т.д. <a href="http://rs.tdwg.org/dwc/terms/measurementValue">http://rs.tdwg.org/dwc/terms/measurementValue</a>
measurementDeterminedBy	В этом поле указываются <i>лица, выполнявшие наблюдения.</i> Имена следует отделять вертикальной чертой, например Oleg Borodin   Alexey Petrovich Seregin <a href="http://rs.tdwg.org/dwc/terms/measurementDeterminedBy">http://rs.tdwg.org/dwc/terms/measurementDeterminedBy</a>
measurementMethod	<i>Ссылка на методику</i> проведения исследований <a href="http://rs.tdwg.org/dwc/terms/measurementMethod">http://rs.tdwg.org/dwc/terms/measurementMethod</a>

## 2.5 Описания, учеты и сборы, выполненные на пробных площадях, маршрутах или трансектах

Данные описаний, учётов и сборов могут быть организованы в виде Sample-Based Data. При оцифровке таких данных информацию необходимо представить в двух таблицах. Первая таблица (Sampling Events) содержит описания самих событий (event) обследования пробных площадей, маршрутов или трансект, проведения уловов и т.п., а вторая (Associated Occurrences) – данные о видах, встреченных в процессе этих событий, с указанием их обилия и иных характеристик.

Для связи двух таблиц используется уникальный для первой таблицы идентификатор eventID, который присваивается не только самим событиям, но и всем относящимся к нему находкам видов во второй таблице. Помимо этого, каждая находка должна иметь собственный, уникальный внутри набора данных идентификатор occurrenceID. Ниже приведен простой пример представления результатов лесной таксации на лесоустроительных площадках, где основной измеряемой величиной является диаметр ствола на уровне груди в см, обозначенный в колонке “organismQuantity” (рис. 10). Фактически вторая таблица составляется в соответствии с правилами Occurrence Data с дополнительным полем EventID.

eventID	eventDate	samplingProtocol
plot1	2019-07-20	Forest inventory FAO (1992)
plot2	2019-07-21	Forest inventory FAO (1992)

eventID	occurrenceID	scientificName	organismQuantity	organismQuantityType
plot1	ISSP-01	Pinus sylvestris L.	85	DBH, cm
plot1	ISSP-02	Picea abies (L.) H.Karst.	60	DBH, cm
plot1	ISSP-03	Betula pendula Roth	30	DBH, cm
plot1	ISSP-04	Populus tremula L.	90	DBH, cm
plot1	ISSP-05	Abies sibirica Ledeb.	40	DBH, cm
plot2	ISSP-06	Pinus sylvestris L.	90	DBH, cm
plot2	ISSP-07	Picea abies (L.) H.Karst.	70	DBH, cm
plot2	ISSP-08	Betula pendula Roth	65	DBH, cm

Рисунок 10. Пример представления данных Sampling events для учетов диаметров деревьев разных видов на пробных площадях.

Обязательные поля таблицы Sampling Events перечислены в табл. 6

Таблица 6. Обязательные поля для представления данных об учете (сборе, описании).

eventID	Уникальный идентификатор описания, учета или сбора. Вы можете составить его самостоятельно, или сгенерировать глобальный идентификатор GUID ( <a href="https://www.guidgenerator.com/">https://www.guidgenerator.com/</a> ). Желательно при составлении ID не ограничиваться сплошной нумерацией записей (1,2,3,4,5 и т.д.), а включить в идентификатор краткое название ООПТ, тип описания или другую подобную информацию. <a href="https://dwc.tdwg.org/terms/#eventID">https://dwc.tdwg.org/terms/#eventID</a>
eventDate	Дата проведения описания, учета или сбора. Должна быть представлена в формате ГГГГ-ММ-ДД, т.е. дата 27 июня 2019 года в DwC будет иметь вид 2019-06-27; временной промежуток 27–29 июня 2019 года – 2019-06-27/29. <a href="https://dwc.tdwg.org/terms/#dwc:eventDate">https://dwc.tdwg.org/terms/#dwc:eventDate</a>
samplingProtocol	Название, ссылка или описание метода, использованного при проведении исследований. <a href="http://rs.tdwg.org/dwc/terms/samplingProtocol">http://rs.tdwg.org/dwc/terms/samplingProtocol</a>

В таблицу Sampling Events также рекомендуется включать данные о географическом положении района исследований, месте размещения постоянных пробных площадей и их размерах (табл. 7).

Таблица 7. Рекомендуемые поля для представления данных об учете (сборе, описании).

sampleSizeValue	Численная характеристика размера пробной площади, маршрута или сбора (продолжительность, длина, площадь, объем). <a href="http://rs.tdwg.org/dwc/terms/sampleSizeValue">http://rs.tdwg.org/dwc/terms/sampleSizeValue</a>
sampleSizeUnit	Единицы измерения величины, указанной в предыдущем поле. <a href="http://rs.tdwg.org/dwc/terms/sampleSizeUnit">http://rs.tdwg.org/dwc/terms/sampleSizeUnit</a>
samplingEffort	Оценка усилий, приложенных для проведения исследований (в свободной форме). Это может время, затраченное на исследование, количество взмахов при окашивании насекомых, число отобранных проб или уловов, число ловушко-суток, километров маршрута и т.п. <a href="http://rs.tdwg.org/dwc/terms/samplingEffort">http://rs.tdwg.org/dwc/terms/samplingEffort</a>
countryCode	Код страны согласно стандарту ISO 3166-1-alpha-2. Его необходимо приводить для однозначного указания страны. Код России - RU. <a href="http://rs.tdwg.org/dwc/terms/countryCode">http://rs.tdwg.org/dwc/terms/countryCode</a>
decimalLatitude	Географическая широта в десятичных градусах. Для северного полушария значения широты положительные, для южного – отрицательные. <a href="http://rs.tdwg.org/dwc/terms/decimalLatitude">http://rs.tdwg.org/dwc/terms/decimalLatitude</a>
decimalLongitude	Географическая долгота в десятичных градусах. Для восточного полушария значения долготы положительные, для западного – отрицательные. <a href="http://rs.tdwg.org/dwc/terms/decimalLongitude">http://rs.tdwg.org/dwc/terms/decimalLongitude</a>
footprintWKT	Координаты находки в формате WKT. Позволяет указывать более сложную пространственную привязку, например, в виде линии или полигона. <a href="http://rs.tdwg.org/dwc/terms/footprintWKT">http://rs.tdwg.org/dwc/terms/footprintWKT</a>
geodeticDatum	Эллипсоид, геодезическая система координат или пространственная система отсчета (SRS), на которой основаны географические координаты. <a href="http://rs.tdwg.org/dwc/terms/geodeticDatum">http://rs.tdwg.org/dwc/terms/geodeticDatum</a>

Обязательные поля таблицы Associated Occurrences перечислены в табл. 8

Таблица 8. Обязательные поля для представления данных о находках видах и их характеристиках, описанных в процессе учета.

eventID	Уникальный идентификатор описания, учета или сбора. Вы можете составить его самостоятельно, или сгенерировать глобальный идентификатор GUID ( <a href="https://www.guidgenerator.com/">https://www.guidgenerator.com/</a> ). Желательно при составлении ID не ограничиваться сплошной нумерацией записей (1,2,3,4,5 и т.д.), а включить в идентификатор краткое название ООПТ, тип описания или другую подобную информацию. <a href="https://dwc.tdwg.org/terms/#eventID">https://dwc.tdwg.org/terms/#eventID</a>
occurrenceID	Уникальный идентификатор находки. Вы можете составить его самостоятельно, или сгенерировать глобальный идентификатор GUID. Желательно при составлении ID не ограничиваться сплошной

	<p>нумерацией записей (1,2,3,4,5 и т.д.), а включить в идентификатор акроним коллекции или аббревиатуру ООПТ, каталожный номер образца или другую подобную информацию.  <a href="https://dwc.tdwg.org/terms/#dwc:occurrenceID">https://dwc.tdwg.org/terms/#dwc:occurrenceID</a></p>
basisOfRecord	<p>В этом поле необходимо указать, <i>что послужило основанием</i> для появления этой записи. Для его заполнения предусмотрен фиксированный набор значений, необходимо использовать одно из них – PreservedSpecimen, FossilSpecimen, LivingSpecimen, MaterialSample, Event, HumanObservation, MachineObservation, Taxon, Occurrence. Практически во всех случаях, описываемых в данном руководстве, это будет <b>HumanObservation</b>, при этом для всех значений важно сохранить оригинальное (слитное) написание.  <a href="https://dwc.tdwg.org/terms/#dwc:basisOfRecord">https://dwc.tdwg.org/terms/#dwc:basisOfRecord</a></p>
eventDate	<p><i>Дата проведения, учета или сбора.</i> Должна быть представлена в формате ГГГГ-ММ-ДД, т.е. дата 27 июня 2019 года в DwC будет иметь вид 2019-06-27; временной промежуток 27-29 июня 2019 года – 2019-06-27/29.  <a href="https://dwc.tdwg.org/terms/#dwc:eventDate">https://dwc.tdwg.org/terms/#dwc:eventDate</a></p>
scientificName	<p><i>Научное название</i> таксона. Заполняется по тем же правилам, что и для таксономических данных. Настоятельно рекомендуется проверить все названия с помощью инструмента Species matching (подробнее см. раздел о представлении таксономических списков)  <a href="https://dwc.tdwg.org/terms/#dwc:scientificName">https://dwc.tdwg.org/terms/#dwc:scientificName</a></p>

В таблице Associated Occurrences важно также представить информацию о присутствии (отсутствии) видов или оценке их обилия (табл. 9).

Таблица 9. Поля DwC для представления данных о присутствии (отсутствии) видов или оценке их обилия в учете.

individualCount	<p><i>Число обнаруженных особей</i>, выраженное в штуках. Например, <b>37</b>.  <a href="http://rs.tdwg.org/dwc/terms/individualCount">http://rs.tdwg.org/dwc/terms/individualCount</a></p>
organismQuantity	<p>Этот термин используется, если для <i>оценки обилия</i> выбрано не число особей, а другая величина. В этом случае значение обилия помещают в этом поле.  <a href="http://rs.tdwg.org/dwc/terms/organismQuantity">http://rs.tdwg.org/dwc/terms/organismQuantity</a></p>
organismQuantityType	<p><i>Размерность величины</i>, указанной в предыдущем поле. Например, <b>r</b> в organismQuantity и <b>BraunBlanquetScale</b> в organismQuantityType.  <a href="http://rs.tdwg.org/dwc/terms/organismQuantityType">http://rs.tdwg.org/dwc/terms/organismQuantityType</a></p>
occurrenceStatus	<p><i>Сведения о присутствии таксона в данном описании, учете или сборе.</i> Может принимать одно из значений – <b>present</b> (таксон присутствует) или <b>absent</b> (таксон отсутствует).  <a href="http://rs.tdwg.org/dwc/terms/occurrenceStatus">http://rs.tdwg.org/dwc/terms/occurrenceStatus</a></p>

Для однозначного сопоставления научных названий с базовой таксономией GBIF желательно также указывать Царство (поле kingdom) и ранг таксона (taxonRank), табл. 10.

Таблица 10. Поля DwC для представления таксономической информации о находках видов, сделанных в процессе учета.

kingdom	<p><i>Царство</i>. Его необходимо указывать для однозначного сопоставления ваших названий с таксономией GBIF. Как известно, существуют виды с одинаковыми названиями и относящиеся к разным таксономическим группам. Кроме того, в названиях ваших таксонов возможны опечатки, или другие неточности. Для заполнения этого поля необходимо указывать линнеевские ранги (“Animalia”), нельзя указывать таксономические группы (“Algae” или “Herbivora”)</p> <p><a href="https://dwc.tdwg.org/terms/#dwc:kingdom">https://dwc.tdwg.org/terms/#dwc:kingdom</a></p>
taxonRank	<p><i>Ранг таксона</i>, до которого удалось определить образец (вид, род, семейство и др.). Это поле необходимо для корректной индексации набора данных на глобальном портале и сопоставления ваших названий с таксономией GBIF. Для заполнения этого поля необходимо использовать линнеевские ранги: царство, тип, класс, порядок, семейство, род, вид. Допускается написание рангов на латинском или английском языке.</p> <p><a href="https://dwc.tdwg.org/terms/#dwc:taxonRank">https://dwc.tdwg.org/terms/#dwc:taxonRank</a></p>

Возможности Darwin Core позволяют не просто представить данные об описаниях, учетах или сборах, но и объединить их в некоторые логические блоки. Например, для работ, выполненных в разные годы на одних и тех же участках, необходимо добавить еще один иерархический уровень – ГОД.

<b>ГОД</b>
УЧЕТНАЯ ПЛОЩАДКА ИЛИ МАРШРУТ
Находки видов на учетной площадке в определенный год

Для представления данных с такой структурой нужен дополнительный идентификатор parentEventID, позволяющий показать принадлежность находки не только к определенной учетной площадке (маршруту), но и к году проведения исследований. Важно не ошибиться в присвоении идентификаторов parentEventID, eventID и occurrenceID. В рассмотренном выше простом примере об измерениях диаметров стволов деревьев на временных площадках схема присвоения ID была следующей:

ПЛОЩАДКА	eventID	
Находки видов на учетной площадке	eventID	occurrenceID

При добавлении года проведения учетов структура данных усложняется и порядок идентификаторов будет следующим:

ГОД	parentEventID		
ПЛОЩАДКА	parentEventID	eventID	
Находки видов на учетной площадке в определенные год	parentEventID	eventID	occurrenceID

Подобным образом можно представить данные почвенно-зоологических исследований, собранных методом разбора почвенных монолитов. Согласно методике, на участке выкапывают несколько почвенных монолитов определенного размера, из которых выбирают всех беспозвоночных, затем определяют их видовую принадлежность, численность и биомассу. Представить данные таких сборов в Darwin Core можно двумя способами – приводя значения обилия беспозвоночных для всего сбора, или для каждого почвенного монолита внутри каждого сбора. В первом случае порядок присвоения ID будет следующим:

СБОР	eventID	
Находки видов в определенном монолите во всем сборе	eventID	occurrenceID

Если приводить информацию по отдельным монолитам внутри сборов, то структура данных усложняется и порядок идентификаторов будет следующим:

СБОР
<b>Почвенный монолит внутри сбора</b>
Находки видов в определенном монолите в определенном сборе

СБОР	parentEventID		
<b>Почвенный монолит внутри сбора</b>	parentEventID	eventID	
Находки видов в определенном монолите в определенном сборе	parentEventID	eventID	occurrenceID

## Раздел 3 ВЕРИФИКАЦИЯ, ИСПРАВЛЕНИЕ И УТОЧНЕНИЕ ДАННЫХ

### 3.1 Основные понятия

После представления данных в формате электронной таблицы необходимо выполнить их проверку и, в случае необходимости, исправление допущенных ошибок и устранение неточностей (data cleaning). Очевидно, что от тщательности проведения

этой процедуры зависят возможности дальнейшего использования данных и надежность выводов, сделанных на основе их анализа. В то же время, усилия по исправлению ошибок в данных должны быть адекватны целям использования этих данных. Следует четко понимать, что ошибки есть всегда, особенно там, где большую роль играет человеческий фактор.

В процессе проверки и исправления важно, чтобы данные не были случайно потеряны, а изменения вносились очень осторожно. Лучше сохранять как старые (исходные) значения, так и новые (исправленные), чтобы при необходимости исходная информация могла быть восстановлена. Крайне желательно документировать все внесенные изменения, указывая сотрудников, которыми эти изменения были сделаны.

Методам проверки, уточнения и оценки качества данных посвящена обширная литература. Основные концепции подробно изложены в работе Chapman (2005). В пособии представлено описание основных типов ошибок и методов их устранения.

### 3.2. Технические ошибки

**Технические ошибки** могут быть связаны с неправильным экспортом или отображением данных, или же их отсутствием в первоисточнике. Чаще всего такие ошибки легко выявить с помощью использования фильтров в программе Excel.

1. *Полнота представления данных.* Есть ли в таблице все необходимые поля (столбцы)? Все ли ячейки заполнены? Желательно не оставлять ячейки без значений пустыми, в них следует указывать “0”, где это осмысленно, либо “нет данных”.
2. *Границы диапазонов.* Если в ваших данных указаны временные или иные диапазоны, необходимо проверить их правильность. Например, указаны ли дни в диапазоне 1–31 (в зависимости от месяца).
3. *Типы данных.* Все ли данные соответствуют полям, в которых они содержатся? Например, в поле month (месяц) могут быть только числовые значения, не текстовые.
4. *Формат данных.* Все ли данные указаны (где это необходимо) в требуемом формате? Например, даты должны быть представлены в формате ГГГГ-ММ-ДД, а координаты – в десятичных градусах ГГ.ГГГГГ.

### 3.3. Ошибки согласованности

**Ошибки согласованности** выявляются применением правил реального мира к данным. Они могут указывать на неправильный ввод данных из более старых записей, например, в гербарных образцах, ошибки транскрибирования (расшифровки этикеток) или последующей обработки. Часто при оцифровке летописей природы можно столкнуться с изменениями в таксономии, в том числе латинских названий видов, изменениями методик сбора материалов и обозначений величин, переименованиями населенных пунктов. Не забывайте, что при подготовке печатного экземпляра

Летописи большую роль играет человеческий фактор, работа по поиску и исправлению опечаток проводится редко. Для поиска и исправления некоторых ошибок согласованности требуется использовать справочные данные, например, список коллекторов. Также следует учитывать стиль работы разных исследователей.

1. *Таксономические и номенклатурные ошибки.* Необходимо проверить корректность написания научных названий. Например, если образцы определены до вида, представлены ли для них биномиальные названия? Кроме того, необходимо проверить, все ли представленные названия корректны.
2. *Ошибки соответствия.* Соответствуют ли даты сбора, идентификации, обновления и оцифровки? Например, нет ли записей, созданных после смерти коллектора (возможно, это другой коллектор с похожим именем)?
3. *Экстремальные значения.* Есть ли выбросы в данных? Являются ли они результатом ошибок или данные корректны?
4. *Географические ошибки.* Рекомендуется визуализировать координаты в ГИС и проверить, все ли точки находок попадают в указанное место или регион. Например, проверить, нет ли находок сухопутных видов в море, или морских видов на суше. Все ли записи с координатами, полученные без применения спутникового навигатора, имеют соответствующую точность и адекватную погрешность?
5. *Неточности в описании методов сборов.* Разные методы исследований (например, учеты на трансектах и пробных площадях) имеют свои особенности. Необходимо проверить, соответствуют ли записи о находках методу сбора данных.
6. *Ошибки последовательности фенологических событий.* Фенологические явления имеют строгую последовательность. Например, цветение предшествует появлению плодов, а фазы цветения (начало, массовое и окончание) имеют только один порядок следования. Начало пожелтения листьев обязательно предшествует массовому листопаду и т.д. При верстке Летописи природы и последующей и оцифровке оператор может допустить опечатку в месяце или дате явления. Зачастую исправление подобных ошибок требует экспертного решения, если не удастся с уверенностью выяснить правильные значения, приходится исключать данные, нарушающие последовательность явлений.

Для выявления и исправления номенклатурных, географических ошибок и выбросов существуют специальные веб-инструменты и методики. Процедура проверки таксономических списков описана в разделе 2.2.

### **3.4. Исправление ошибок после публикации данных в GBIF.org**

После того, как набор данных опубликован в сети GBIF, авторы могут вносить в него исправления, а также добавлять новые записи или удалять некорректные, если

таковые будут выявлены. Для обновления набора данных необходимо подготовить таблицу с актуальными стандартизированными данными и войти в уже опубликованный набор данных в IPT. Затем нужно удалить исходный файл с данными, загрузить исправленные (обновленные) данные и соотнести заголовки полей обновленной таблицы с терминами Darwin Core. При необходимости можно отредактировать метаданные. После нажатия кнопки *Publish*, через портал GBIF будет доступна новая версия набора данных.

Очень важно при подготовке очередной версии набора данных сохранять исходные идентификаторы для записей (из первой опубликованной версии). IPT хранит все версии набора данных и, если авторы при публикации очередной версии присваивают новые идентификаторы уже опубликованным записям, эти записи будут восприниматься системой как вновь публикуемые.

## Раздел 4. МАТЕРИАЛЫ ДЛЯ ПРАКТИЧЕСКИХ ЗАНЯТИЙ НА КУРСЕ ПОВЫШЕНИЯ КВАЛИФИКАЦИИ

Таблицы для выполнения упражнений доступны через [гугл-диск](#)



фото Наталья Иванова CC-BY

*Эта вымышленная история основана на реальных данных и используется исключительно для учебных целей. Концепция некоторых упражнений позаимствована из курсов повышения квалификации международного проекта BioDATA.*

### Описание ситуации

Заповедник “Вепревы Топи” находится в живописной части Вологодской области, в 50 км северо-западнее города Великий Гусляр в междуречье Юга и Лузы. Заповедник создан для сохранения уникальных участков старовозрастных хвойно-широколиственных лесов. На его территории известны 807 видов растений и 545 видов животных.

С момента создания заповедника в 1999 году силами научного отдела проводятся исследования по программе “Летопись природы”. На данный момент все тома Летописи существуют только в бумажном виде. Ежегодно в заповеднике проводятся научные исследования по изучению биоразнообразия, выполняемые сотрудниками ведущих российских научных центров. Отчеты о результатах этих работ хранятся в научном отделе заповедника. В музее природы заповедника хранятся гербарий и энтомологическая коллекция.

На сегодняшний день в виде электронных таблиц представлены только Кадастр биологического разнообразия заповедника, данные о фенологических наблюдениях, а также материалы по оценке урожайности ягодников за текущий год. Коллекции заповедника не оцифрованы и не имеют электронного каталога. С этого года заповедник участвует в международном проекте “Летопись природы Евразии” и начал

работу по оцифровке данных Летописей и других материалов о биоразнообразии, чтобы сделать их доступными для исследователей всего мира.

### Упражнение 1. Проверка и стандартизация списка видов заповедника

Было принято решение на первом этапе оцифровки стандартизировать список млекопитающих заповедника. Вам поручено проверить данные на соответствие таксономической базе GBIF и при необходимости исправить ошибки.

Используйте файл **ChecklistDataExample.xlsx**, шаблон для представления таксономических данных из файла **ChecklistTemplate.xlsx** и веб-инструмент для проверки таксономии <https://www.gbif.org/tools/species-lookup>

1. Объедините родовое и видовое названия, с помощью формул EXCEL
2. Выполните проверку таксономии, при необходимости исправьте названия таксонов онлайн, загрузите результаты в виде CSV файла на свой компьютер.
3. Представьте данные в таблице с полями DwC (используя шаблон). Используйте в качестве taxonID идентификаторы таксонов GBIF Backbone.

Используйте для справки материалы раздела 2.2 данного пособия.

### Упражнение 2. Оцифровка этикеточных данных Гербария заповедника

На данный момент руководство ООПТ не имеет возможности массового сканирования гербарных листов, хранящихся в музее природы заповедника. Тем не менее, очевидна необходимость создания электронного каталога коллекции, включающего максимально полную этикеточную информацию. Для организации данных было решено использовать стандарт Darwin Core. Используя изображения из архива **HerbariumScans.zip** и шаблон **TemplateHerbarium.xlsx**, перенесите информацию с этикеток в электронную таблицу. Какие поля DwC вы могли бы добавить для более полного представления этикеточных данных?

Оригинальное описание терминов Darwin Core (на английском языке) доступно на официальном сайте <https://dwc.tdwg.org/terms/>, краткое описание на русском – [http://gbif.ru/DwC spec](http://gbif.ru/DwC_spec). Также справочные материалы доступны в разделе 2.3 данного пособия.

### Упражнение 3. Sampling Events – простой способ представления данных об урожайности ягодников в Darwin Core

В заповеднике ежегодно проводятся исследования по оценке урожайности ягодников. Эти учеты выполняются на площадках площадью 100 м<sup>2</sup> по стандартной методике, описанной в ГОСТ-17268-71. Данные, собранные в текущем полевом сезоне были занесены в электронную таблицу **SamplingEventDataExample.xlsx**. Ознакомьтесь с таблицей и определите, какие данные описывают учетные площадки, а

какие – находки видов на площадках. Заполните таблицу ниже. Используя шаблон в файле **TemplateSamplingEvent.xlsx** и перечень обязательных для заполнения полей, приведите данные об урожайности ягод на учетных площадках к стандарту Darwin Core. Вся ли необходимая для этого информация представлена в таблице с исходными данными?

Таблица для выполнения упражнения 3

Поля, характеризующие учетные площадки	Поля, характеризующие виды на учетных площадках
Число записей в таблице <b>Sampling Events</b> (Число учетных площадок)	Число записей в таблице <b>Associated Occurrences</b> (Общее число находок видов)

Информацию об обязательных и рекомендуемых для заполнения полях см. в таблицах 6–9 раздела 2.5.

### Упражнение 3а. **Sampling Events** – продвинутый способ представления данных: сборы дождевых червей методом разбора почвенных проб (домашнее задание для желающих)

Внимательно прочитайте текст задания и подумайте, чем структура этих данных отличается от структуры данных об урожайности ягодников, собранных в разные годы на одних и тех же площадках?

В окрестностях заповедника изучали разнообразие дождевых червей на бывших сельскохозяйственных землях. Было исследовано три биотопа: луг (залежь), молодой березняк на месте пашни и старый смешанный лес с участием широколиственных деревьев. Исследования проводили методом разбора почвенных проб. В каждом биотопе было разобрано по 8 почвенных монолитов размером 25×25×25 см. Червей фиксировали в 4%-м растворе формалина и определяли в лабораторных условиях. Собранные данные доступны в файле **SamplingEventDataExamplehomework.xlsx**. Вам поручено представить эти данные в стандарте Darwin Core, отразив не только принадлежность находок к сборам в разных биотопах, но и разнообразие видов, найденных внутри каждой пробы. Для работы используйте шаблон **TemplateSamplingEvent-homework.xlsx**. Вы можете прислать заполненные таблицы на проверку преподавателям курса на электронный адрес: [gbif.ru@yandex.ru](mailto:gbif.ru@yandex.ru).

#### Упражнение 4. Верификация и исправление данных

В окрестностях заповедника в июне 2010 г. проводили исследования разнообразия панцирных клещей. Данные были перенесены из бумажных носителей в электронную таблицу **DataCleaningExample.xlsx**. Используя контрольный список для проверки и таблицу ниже, найдите и исправьте ошибки в данных. Задokumentируйте свои действия в таблице.

Контрольный список для проверки данных:

- Все названия таксонов написаны правильно
- Ранги всех таксонов указаны корректно
- Все записи имеют координаты
- Все точки находок находятся в междуречье Юга и Лузы
- Все координаты преобразованы в десятичные градусы
- Все даты находятся в соответствующем столбце и представлены в формате ГГГГ-ММ-ДД

Таблица. Предлагаемые действия по поиску и исправлению ошибок в данных

Столбец	Подсказка	Ошибка	Какое исправление сделано	Кто внес исправление
eventDate	Проверьте наличие пустых значений			
YE	Данные собраны в 2010-м году			
countryCode	Проверьте отсутствие пустых значений			
decimalLatitude	Все ли точки сборов находятся в северном полушарии?			
decimalLongitude	Все ли координаты указаны в десятичных градусах?			
coordinateUncertainty	Значение следует указывать в метрах			
taxonRank	Для всех ли таксонов ранги указаны корректно?			
species	Все ли названия соответствуют таксономической базе GBIF? Используйте онлайн инструмент <a href="https://www.gbif.org/tools/specieslookup">https://www.gbif.org/tools/specieslookup</a>			



### Список литературы

*Филонов К.П., Нухимовская Ю.Д.*, 1990. Летопись природы в заповедниках СССР. Методическое пособие. М.: Наука. 143с.

*Шашков М.П., Иванова Н.В.*, 2018 Стандарты и веб-инструменты для представления данных через глобальные порталы по биоразнообразию // Доклады Международной конференции «Математическая биология и биоинформатика». Под ред. В.Д. Лахно. Том 7. Пушино: ИМПБ РАН. Статья № е98. DOI: 10.17537/icmbb18.55

*Шашков М.П. и др.*, 2018. Возможности, решения и инструменты GBIF для оцифровки и развития естественнонаучных коллекций // Зоологические исследования. №20. С. 169–174.

*Шашков М.П., Чадин И.Ф., Иванова Н.В.*, 2017. Методические рекомендации по стандартизации данных для публикации через глобальный портал GBIF.org и подготовке статьи о данных // Труды Кольского научного центра РАН. В. 5. С. 22–35.

*Bisby F.A.*, 2000. The quiet revolution: Biodiversity informatics and the Internet // Science. V. 289. P. 2309–2312. DOI: 10.1126/science.289.5488.2309

*Chapman A.D.*, 2005. Principles of Data Quality. Version 1.0. Report for the Global Biodiversity Information Facility, Copenhagen. DOI: 10.15468/doc.jrgg-a190.

*Robertson T.*, et al., 2014. The GBIF Integrated Publishing Toolkit: Facilitating the Efficient Publishing of Biodiversity Data on the Internet. PLoS ONE 9(8): e102623. DOI: 10.1371/journal.pone.0102623

*Wheeler Q.D.*, 2004. What if GBIF? // Bioscience. Vol. 54. I. 8. P. 717. DOI: 10.1641/0006-3568(2004)054[0718:WIG]2.0.CO;2

*Wieczorek J.* et al., 2012. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard. PLoS ONE 7(1): e29715. DOI: 10.1371/journal.pone.0029715

*Wilkinson M.D.* et al., 2016. The FAIR Guiding Principles for scientific data management and stewardship // Scientific Data. V. 3. Article number: 160018. DOI: 10.1038/sdata.2016.18