

Государственное научно-производственное объединение
«Научно-практический центр НАН Беларуси по биоресурсам» GBIF Republic of Belarus

КРАТКОЕ РУКОВОДСТВО

ПО ПОДГОТОВКЕ НАБОРОВ ДАННЫХ
ДЛЯ ПУБЛИКАЦИИ ЧЕРЕЗ ГЛОБАЛЬНЫЙ
ИНФОРМАЦИОННЫЙ ФОНД
ПО БИОРАЗНООБРАЗИЮ (GBIF)

Минск
«Экоперспектива»
2022

УДК 574(476):004(083.13)
ББК 32.973
К78

Составитель О.И. Бородин

Краткое руководство по подготовке наборов данных для публикации через Глобальный информационный фонд по биоразнообразию (GBIF) / сост. : О.И. Бородин. — Минск : Экоперспектива, 2022. — 64 с.

ISBN 978-985-469-896-0.

Руководство содержит общую информацию о Глобальном информационном фонде по биоразнообразию (GBIF), а также алгоритм создания и внесения в него наборов данных. Создано в рамках проекта «Укрепление потенциала на основе регионального сотрудничества в целях мобилизации данных в Беларуси, Латвии и Литве / Capacity Building through Regional Cooperation for Data Mobilization in Belarus, Latvia and Lithuania (CESP2020-015)» (Программа поддержки укрепления потенциала GBIF (CESP) / Capacity Enhancement Support Program GBIF (CESP)).

УДК 574(476):004(083.13)
ББК 32.973

ISBN 978-985-469-896-0

© Бородин О.И., составление, 2022
© Оформление. УП «Экоперспектива», 2022

ВВЕДЕНИЕ

Глобальная база данных по биоразнообразию (Global Biodiversity Information Facility (GBIF)) является международной открытой системой обмена данными о биологическом разнообразии. Система финансируется правительствами стран-участниц. GBIF позволяет любому пользователю получить доступ к информации, находящейся в системе.

В настоящее время в базе данных содержится более 2,13 млрд записей, что делает эту систему крупнейшим в мире ресурсом по биологическому разнообразию. Эти данные получены из самых разных источников, начиная от музейных образцов, опубликованных данных, научных отчетов, диссертаций и заканчивая фотографиями со смартфонов, сопровождаемые данными о дате и времени наблюдения и геотегами, которые через интегрированные с GBIF ресурсы (например, iNaturalist) помещают любители-натуралисты.

Деятельность GBIF координируется через Секретариат, располагающийся в Копенгагене, и обеспечивается функционированием узлов (Nodes), которые в свою очередь координируются Национальными представительствами и объединяют организации (научно-исследовательские институты, музеи, ботанические сады и т. д.), публикующие данные о биоразнообразии.

Следует особенно отметить, что опубликовать свои данные в GBIF может любой желающий, но сделать он это может только через зарегистрированные в системе публикующие организации. Таким образом возрастает ответственность за качество и достоверность публикуемых материалов.

В настоящее время членами GBIF является 104 участника, представленные 63 странами и рядом крупных, международных организаций.

GBIF взаимодействует с большим количеством природоохранных инициатив. В качестве примера можно указать Конвенцию о биологическом разнообразии, Конвенцию по мигрирующим видам, Энциклопедия жизни (EOL), Партнерство по индикаторам биоразнообразия (BIP), Сохранение арктической флоры и фауны (CAFF), DataONE, Глобального партнерства по сохранению растений (GPPC), Информационной системой о биологическом разнообразии океанов (OBIS) и др.

GBIF объединяет все эти источники информации посредством использования стандартов данных, таких как Darwin Core, который составляет базис для основной части записей о встречаемости видов.

Авторы данных предоставляют открытый доступ к своим наборам данных, используя машиночитаемые обозначения лицензий Creative Commons, что позволяет ученым, исследователям и другим лицам применять данные в сотнях рецензируемых публикаций и политических документов каждый год.

Каждый из этих нюансов будет более подробно рассмотрен в соответствующих разделах руководства.

Следует подчеркнуть, что данное руководство представляет собой собрание базовых методов подготовки наборов данных для их публикации через GBIF. Здесь не рассматривается теория информатики о биоразнообразии, нет обсуждения теории управления данными и т. д. Возможно в перспективе будет подготовлено более обширное издание, охватывающее все аспекты информатики о биоразнообразии в приложении к GBIF. На данном этапе при необходимости можно воспользоваться многочисленными руководствами доступными на сайте gbif.org и ассоциированных ресурсах.

1. Общий алгоритм подготовки наборов данных для их публикации через GBIF

Существуют различные алгоритмы публикации наборов данных в GBIF, начиная от использования специализированного программного обеспечения, позволяющего автоматически, в том числе на регулярной основе переводить существующую в формате различных баз данных информацию в адаптированную для GBIF форму.

Как уже подчеркивалось, в рамках данного руководства мы рассматриваем базовый вариант подготовки набора данных, который позволит пользователям получить общее представление об этом процессе, а по мере наработки опыта и компетенции со временем использовать все возможности сообщества GBIF.

Ниже в рамках данного краткого руководства рассмотрены следующие вопросы, имеющие отношение к подготовке набора данных для его публикации в GBIF.

Для полноценной работы с системой GBIF первым шагом является регистрация на сайте gbif.org. Этот алгоритм рассмотрен в разделе 2. Далее, как уже было отмечено в вводной части, публикация данных в GBIF осуществляется любым пользователем, но от имени конкретной организации, получающей статус публикующей организации, с сохранением авторства создателя набора данных. Эти моменты рассмотрены в разделе 3, посвященном алгоритму регистрации организации в качестве публикующей, и разделе 8, где подчеркивается, на каких этапах обозначаются права на датасет.

В самом упрощенном варианте набор данных перед загрузкой в GBIF должен быть преобразован в табличную форму. Столбцы в данной таблице озаглавлены строго определенным образом. Каждое заглавие столбца должно соответствовать определенному термину стандарта Darwin Core. Как выглядят таблицы, какие существуют шаблоны и какие типы данных используются в GBIF рассматривается в разделе 5.

Словарь используемых терминов Darwin Core с их кратким описанием приведен в разделе 4.

После формирования таблицы в качестве первичного набора данных рекомендуется провести дополнительную оценку качества информации. Наряду с разработкой алгоритмов получения высококачественной информации, которые являются предметом отдельного методического пособия, в разделе 6 приводятся примеры существующих инструментариев, при помощи которых можно оптимизировать очистку полученного набора данных от теоретических ошибок.

В разделе 7 рассматриваются вопросы лицензирования публикуемых через GBIF наборов данных и рекомендации на этот счет со стороны сообщества GBIF.

Наконец в разделе 8 приводится пошаговый обзор загрузки созданного набора данных через IPT, особый инструментарий, являющийся наиболее распространенным вариантом при загрузке наборов данных через GBIF. В этом же разделе рассматриваются процессы описания (подготовки метаданных) загружаемого в систему набора данных. В качестве примера выбрана тестовая версия набора данных по цикадовым Беларуси, Латвии и Литвы.

В разделе 9 даны краткие указания о возможностях использования всего разнообразия IPT, через которые в настоящее время может быть загружена информация о биоразнообразии в GBIF в мире.

Завершает руководство перечень рекомендуемых источников, в которых можно получить более детальную информацию по отдельным вопросам подготовки наборов данных, обеспечения их высокого качества и загрузки в GBIF.

2. Регистрация на сайте в качестве пользователя gbif.org

Для получения доступа к подавляющему большинству функций портала, в том числе возможности скачивания сформированных запросов с верным алгоритмом их цитирования, необходимо пройти регистрацию.

Для этого следует нажать кнопку «Login» в правом верхнем углу главной страницы (рисунок 1).

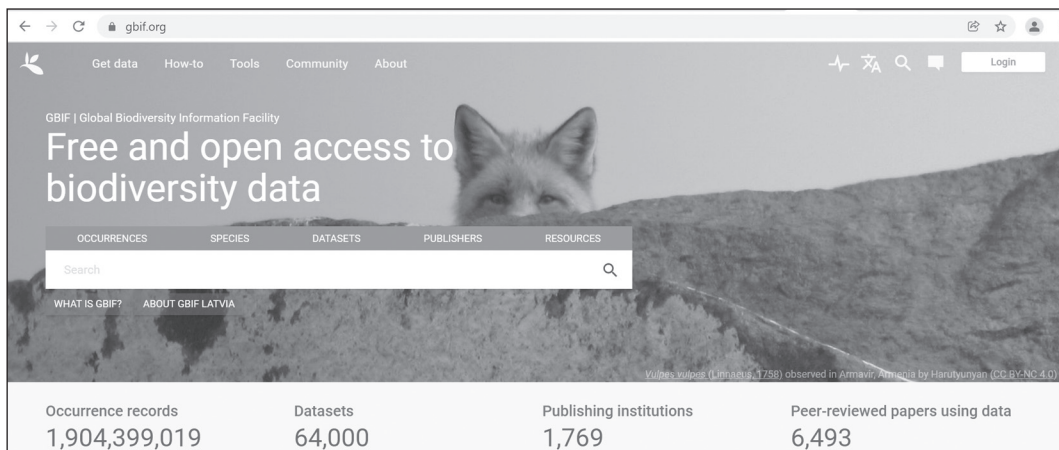


Рисунок 1

Появится окно для авторизации (рисунок 2). В том случае, если вы уже авторизованы на сайте, на данном этапе можно войти в существующий аккаунт. Для регистрации необходимо перейти на закладку «Register» (рисунок 3). После заполнения формы необходимо пройти соответствующий тест CAPTCHA (рисунок 4).

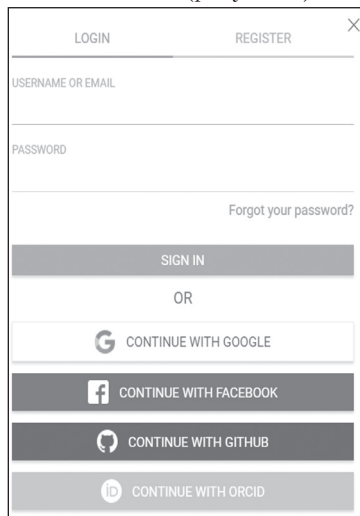


Рисунок 2

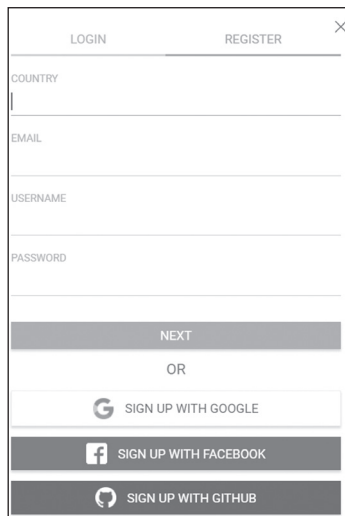


Рисунок 3



Рисунок 4

После нажатия кнопки «Sign up» появляется сообщение о необходимости продолжения регистрации с использованием ссылки, которая выслана на e-mail, указанный в регистрационной форме (рисунок 5).

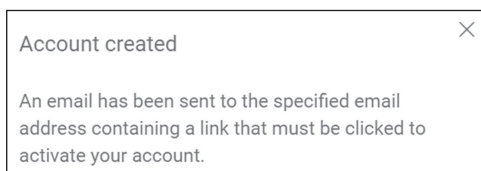


Рисунок 5

На e-mail будет получено сообщение следующего содержания (рисунок 6):

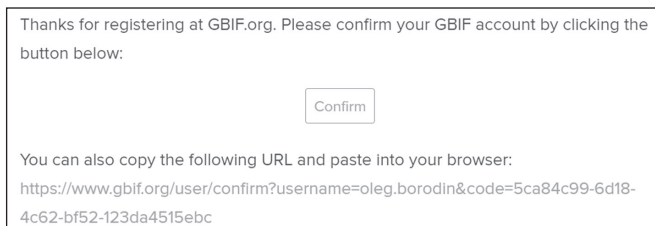


Рисунок 6

После перехода по ссылке либо нажатия на кнопку «Confirm» произойдет возврат на сайт gbif.org и появится форма верификации (рисунок 7).

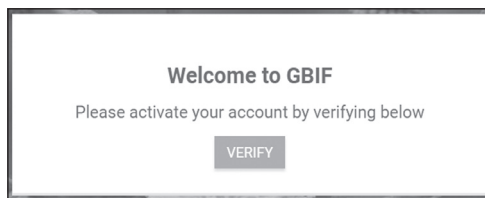


Рисунок 7

После нажатия кнопки «Verify» появляется сообщение о необходимости подтверждения понимания Политики конфиденциальности GBIF (рисунок 8). Более подробную информацию по данному вопросу можно посмотреть по адресу: <https://www.gbif.org/terms/privacy-policy>.

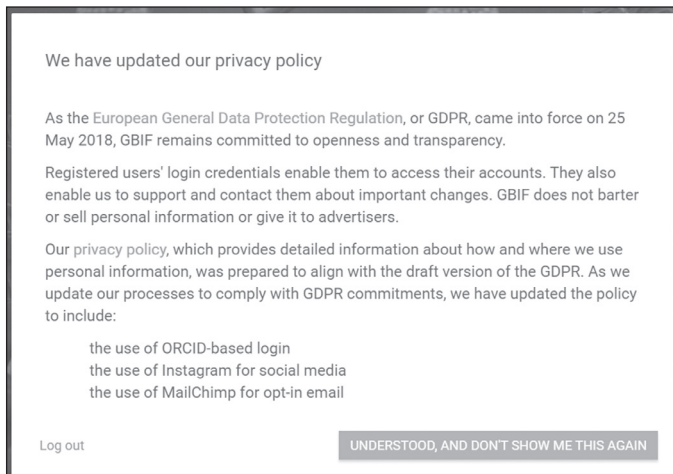


Рисунок 8

После нажатия кнопки «Understood, ...» появляется сообщение о создании вашего аккаунта (рисунок 9). Нажатием кнопки «Continue» вы завершаете регистрацию.

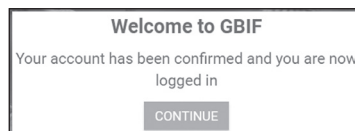


Рисунок 9

3. Регистрация организации в качестве публикующей

Согласно существующим правилам в GBIF публиковать данные можно только от имени организаций (учреждений, сообществ, сетей и т. д.), не от имени частных лиц. Сам алгоритм загрузки в систему GBIF наборов данных предполагает его связь с конкретной организацией.

В связи с этим необходимо убедиться в том, имеет ли организация, от имени которой планируется осуществлять публикацию данных, статус публикующей в GBIF.

Сделать это можно на сайте gbif.org несколькими способами, например, в разделе «Community», где необходимо перейти по ссылке в раздел «Publishers» (рисунок 10).

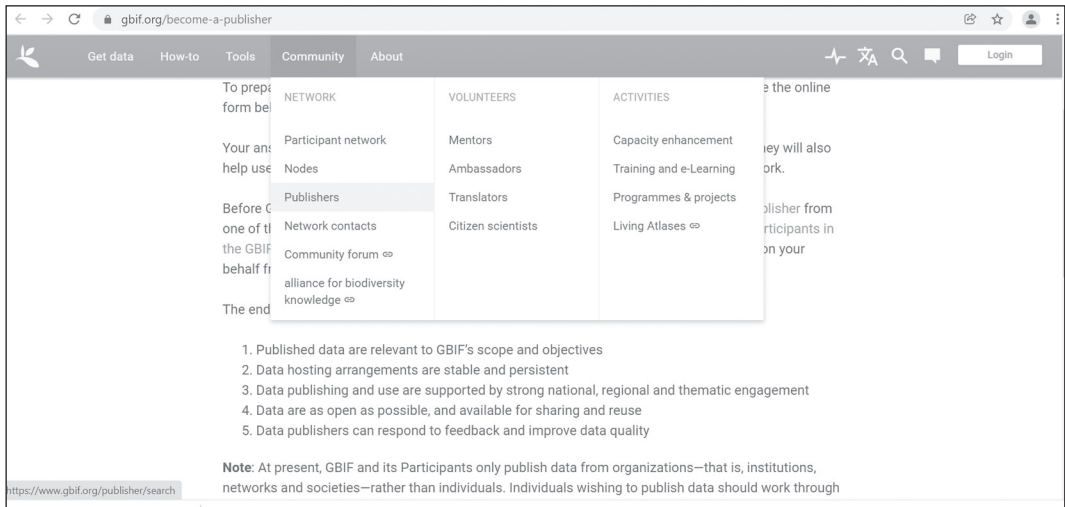


Рисунок 10

На момент подготовки руководства в системе GBIF было зарегистрировано более 2100 организаций (рисунок 11).

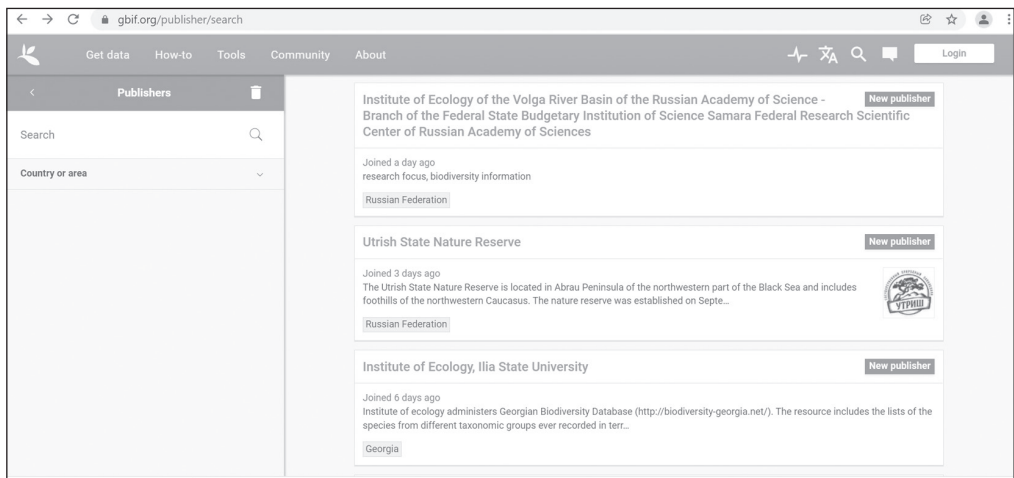


Рисунок 11

В данном списке отображается в том числе информация о времени, когда та или иная организация присоединилась к деятельности GBIF. Здесь же предусмотрен и вариант поиска организации по названию либо по региону (рисунок 12).

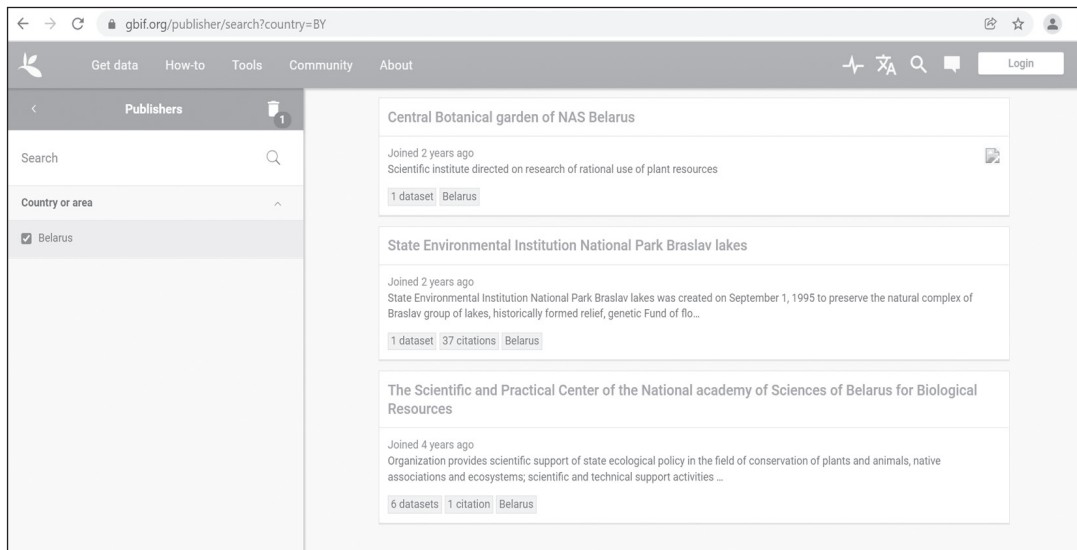


Рисунок 12

Если у организации еще нет статуса публикующей, сделать это можно, заполнив соответствующую онлайн-форму (см. ниже). Безусловно, предварительно необходимо получить согласие со стороны руководства организации о присоединении к системе GBIF. Каких-либо документальных подтверждений для GBIF не требуется. Формы подтверждения такого рода согласия остаются на усмотрение организаций.

Алгоритм получения организацией статуса публикующей подробно описан в разделе «Become to publisher» (рисунок 13).

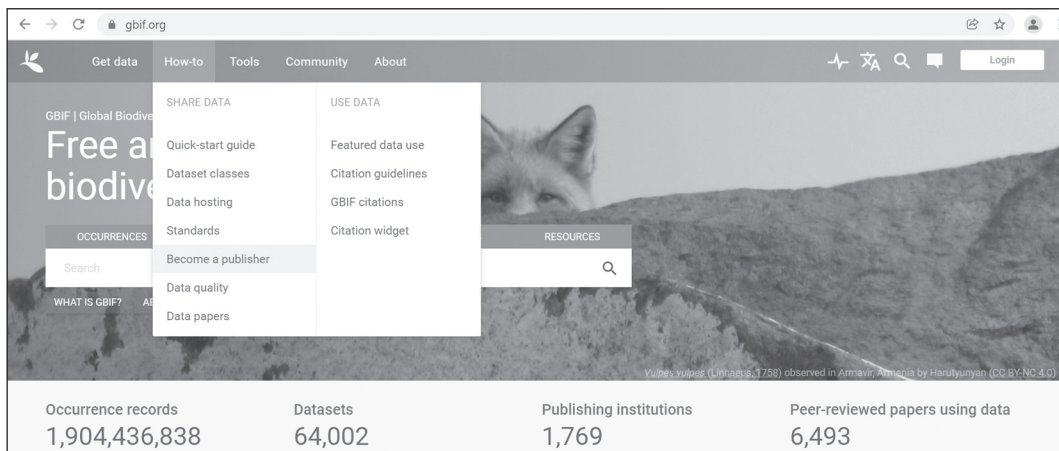
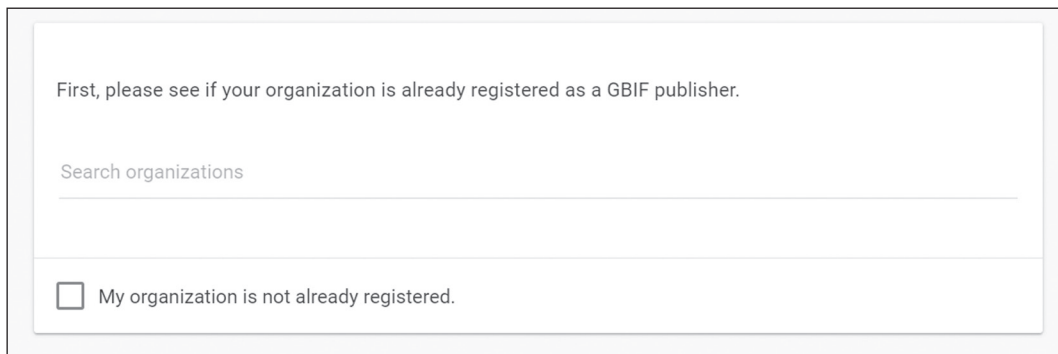


Рисунок 13

В нижней части данного раздела также присутствует поисковая форма, позволяющая найти необходимую организацию. Если организация не имеет статус публикующей, следует отметить пункт «My organization is not already registered», подтверждающий это (рисунок 14).



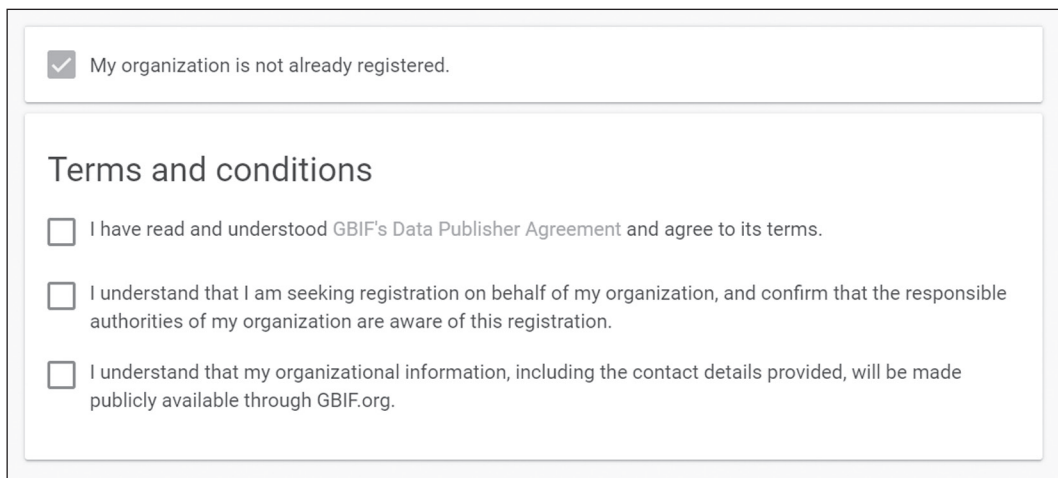
First, please see if your organization is already registered as a GBIF publisher.

Search organizations

My organization is not already registered.

Рисунок 14

В появившемся диалоговом окне необходимо подтвердить, что принимается Соглашение о публикующей в GBIF организации; подтверждается, что ответственные лица в организации осведомлены о происходящем присоединении к GBIF и представленные при регистрации данные могут быть общедоступными (рисунок 15).



My organization is not already registered.

Terms and conditions

I have read and understood GBIF's Data Publisher Agreement and agree to its terms.

I understand that I am seeking registration on behalf of my organization, and confirm that the responsible authorities of my organization are aware of this registration.

I understand that my organizational information, including the contact details provided, will be made publicly available through GBIF.org.

Рисунок 15

После того, как все эти пункты будут отмечены, появится онлайн-форма, в которой будет предложено внести краткую информацию об организации, включая контактные данные, сферу деятельности (рисунок 16). Информацию можно вносить на разных языках, но английская версия должна быть в любом случае.

Organization details

Your chance to tell GBIF users who you are and what you do.

Organization name *		Home page	
Email		Phone	
Organization email e.g. secretariat@fibg-museum.org			
Address *			
City *	Province	Postal code	
Country *	Logo E.g. http://my.organization.org/images/logo.png		
Description *			

In *English*, please briefly describe the scope of your institution/organization in relation to GBIF's mission (e.g. collection holdings, research focus, biodiversity information management, etc.). This description will appear on your publisher page. You may also wish to include a version in another language, but English is required.

Рисунок 16

Информация, представленная на данном этапе, в конечном итоге будет отображаться на персональной странице организации после ее одобрения в качестве публикующей.

Процедура одобрения (Endorsement) предполагает 3 варианта. В случае, если страна уже присоединилась к системе GBIF, в форме одобрения будет автоматически предложен Национальный узел, через который может проходить процесс одобрения (рисунок 17).

Endorsing node

To support publishers and review data quality all publishers are associated with a GBIF node. Please check the suggestion below, and correct it if needed:

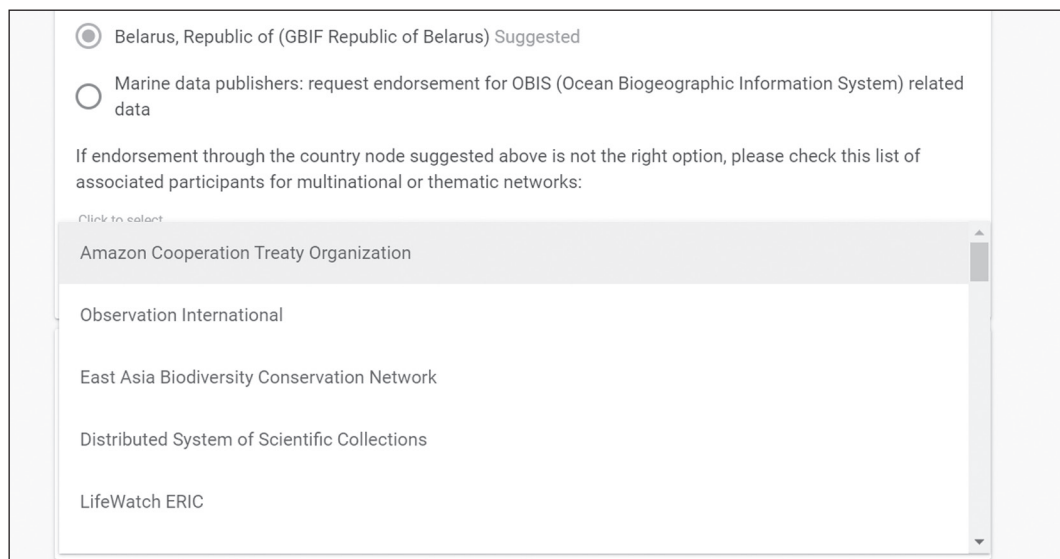
- Belarus, Republic of (GBIF Republic of Belarus) Suggested
- Marine data publishers: request endorsement for OBIS (Ocean Biogeographic Information System) related data

If endorsement through the country node suggested above is not the right option, please check this list of associated participants for multinational or thematic networks:

Click to select

Рисунок 17

Для организаций, работающих в области морской биологии, по умолчанию также предлагается вариант одобрения через OBIS (Информационную систему биоразнообразия океана). Также можно выбрать других партнеров GBIF из довольно обширного выпадающего списка (рисунок 18).



The screenshot shows the 'Endorsing node' form with the following content:

- Belarus, Republic of (GBIF Republic of Belarus) Suggested
- Marine data publishers: request endorsement for OBIS (Ocean Biogeographic Information System) related data

If endorsement through the country node suggested above is not the right option, please check this list of associated participants for multinational or thematic networks:

Click to select

- Amazon Cooperation Treaty Organization
- Observation International
- East Asia Biodiversity Conservation Network
- Distributed System of Scientific Collections
- LifeWatch ERIC

Рисунок 18

Если ни один из предложенных вариантов не подходит, то можно выбрать пункт «Help me with endorsement» (рисунок 19).

Endorsing node

To support publishers and review data quality all publishers are associated with a GBIF node. Please check the suggestion below, and correct it if needed:

Help me with endorsement
 Marine data publishers: request endorsement for OBIS (Ocean Biogeographic Information System) related data

If endorsement through the country node suggested above is not the right option, please check this list of associated participants for multinational or thematic networks:

Click to select ▼

Рисунок 19

Также предусмотрены варианты, в которых можно указать номер проекта GBIF, если в нем участвует организация в качестве одного из исполнителей.

На следующем этапе необходимо указать данные ответственного лица, выступающего в качестве контактной точки (рисунок 20). Здесь же можно указать и дополнительные контактные лица, включая IT-специалиста.

Contacts

We need to know how to keep in touch with you.

First name * Last name *

Email * Phone
Remember to prefix with country code

People move on! Please add at least one alternate contact, and consider using a generic email e.g. helpdesk@a.com that will always reach an appropriate person.

Add administrative contact Add technical contact

Рисунок 20

На указанные адреса будут посланы, наряду с подтверждающим регистрацию приветственным письмом, учетные данные, которые понадобятся в последующем на этапе публикации данных от имени организации через IPT (The Integrated Publishing Toolkit), о чем более подробно будет сказано в разделе 8.

На следующем этапе будет предложено указать, какого рода данные планируется публиковать от имени организации (рисунок 21). Данный вопрос более подробно рассматривается в разделе 5 руководства.

The screenshot shows a form titled "What and how". The text reads: "Help us understand what kind of data you plan to publish, and what support you may need." Below this, it states: "GBIF.org supports publication of four types of data, explained here. Responsibility for formatting the data and hosting the original datasets remains with the data publisher, but we can help you find appropriate technical solutions." There are four checkboxes: "Resources metadata", "Checklist data", "Occurrence-only data", and "Sampling-event data". A text input field is labeled "Data description *". At the bottom, a smaller text input field is labeled "What kinds of relevant data do you have that you intend to publish through GBIF? Please give a brief description."

Рисунок 21

По завершению заполнения онлайн-формы (рисунок 22) необходимо указать, каким образом планируется осуществлять публикацию данных: через уже существующие IPT (см. раздел 9) либо через IPT, который планируется установить на сервере организации. Также предлагается указать, требуется ли помощь в публикации данных.

The screenshot shows three radio button questions: "Do you have EITHER the capacity to run a live server, OR access to a server, through which you will make your original dataset available to GBIF.org?", "Are you planning to install and run publishing software (such as the Integrated Publishing Toolkit – IPT to publish your data directly to GBIF.org)?", and "Do you need help in publishing your data?". At the bottom right, there is a button labeled "REGISTER ORGANIZATION".

Рисунок 22

После заполнения всех разделов формы необходимо нажать кнопку регистрации и отправить ее на обработку в Секретариат GBIF.

4. Стандартизация наборов данных

Для унификации и стандартизации собираемой во всем мире информации о биологическом разнообразии в GBIF используются открытые и общепризнанные стандарты, основным из которых является стандарт Darwin Core, который целенаправленно был разработан для хранения данных о биоразнообразии.

В частности, словарь таких терминов доступен по ссылке <https://dwc.tdwg.org/list/>, где приводится их более подробное описание с приведением примеров использования и комментариями.

Ниже приводится список основных терминов, рубрифицированных по разделам, с краткой информацией о каждом из них.

1. Record-level

Категория содержит универсальные термины, которые могут применяться к записям любого типа в наборе данных.

1.1. type — типология ресурса.

1.2. modified — дата и время последнего изменения ресурса.

1.3. language — язык ресурса.

1.4. license — юридический документ, дающий официальное разрешение осуществления деятельности с ресурсом.

1.5. rightsHolder — лицо или организация, владеющие правами на ресурс или управляющие им.

1.6. accessRights — информация о том, у кого есть доступ к данному ресурсу или о его статусе защиты.

1.7. bibliographicCitation — библиографическая ссылка на ресурс.

1.8. references — связанный ресурс, на который ссылается либо цитирует или каким-либо иным образом указывает описанный ресурс.

1.9. institutionID — идентификатор учреждения, в котором находится объект (объекты) или информация, указанная в записи.

1.10. collectionID — идентификатор коллекции или набора данных, из которого была получена запись.

1.11. datasetID — идентификатор набора данных, может быть глобальным уникальным идентификатором или идентификатором, специфичным для коллекции или учреждения.

1.12. institutionCode — наименование (или аббревиатура), используемое учреждением, хранящим объект (объекты) или информацию, указанную в записи.

1.13. collectionCode — название, акроним, шифр или аббревиатура, идентифицирующие коллекцию или набор данных, из которого была получена запись.

1.14. datasetName — название, определяющее набор данных, из которого была получена запись.

1.15. ownerInstitutionCode — название (или акроним), используемое учреждением, имеющим право собственности на объект (объекты), или информация, указанная в записи.

1.16. basisOfRecord — специфика записи данных.

1.17. informationWithheld — дополнительная информация, которая не была опубликована в данной записи.

1.18. dataGeneralizations — действия, предпринятые для того, чтобы сделать общие данные менее конкретными или полными, чем в исходной форме, предполагает, что альтернативные данные более высокого качества могут быть доступны по запросу.

1.19. dynamicProperties — список дополнительных измерений, фактов, характеристик или утверждений о записи.

2. Occurrence

Категория описывает существование организма (в понимании <http://rs.tdwg.org/dwc/terms/Organism>) в определенном месте в определенное время.

2.1. occurrenceID — идентификатор наблюдения (в отличие от определенной цифровой записи наблюдения), при отсутствии постоянного глобального уникального идентификатора создайте одну из комбинаций идентификаторов в записи, которые наиболее точно определяют идентификатор occurrenceID глобально уникальным.

2.2. catalogNumber — идентификатор (предпочтительно уникальный) для записи в наборе данных или коллекции.

2.3. recordNumber — идентификатор, заданный для наблюдения во время записи, часто служит связующим звеном между заметками полей и записью наблюдения, например, номером сборщика образцов.

2.4. recordedBy — список (объединенный и раздельный) имен людей, групп или организаций, ответственных за запись оригинального наблюдения, первым должен быть указан основной сборщик или наблюдатель, особенно тот, который применяет личный идентификатор (recordNumber).

2.5. recordedByID — список (объединенный и раздельный) глобального уникального идентификатора для лица, людей, групп или организаций, ответственных за запись оригинального наблюдения.

2.6. individualCount — количество особей, присутствовавших в момент наблюдения.

2.7. organismQuantity — числовое или исчисляемое значение количества организмов.

2.8. organismQuantityType — тип количественной системы, используемой для определения количества организмов.

2.9. sex — пол биологического индивида(ов), представленного в наблюдении.

2.10. lifeStage — возрастной класс или жизненная стадия организма(ов) на момент регистрации наблюдения.

2.11. reproductiveCondition — репродуктивное состояние биологического индивида(ов) во время наблюдения.

2.12. behavior — поведение, показанное субъектом на момент записи наблюдения.

2.13. establishmentMeans — заявление о том, был ли организм или организмы введены в данное место и время посредством прямой или косвенной современной активности человека.

2.14. degreeOfEstablishment — степень, в которой организм выживает, воспроизводит и расширяет свой диапазон в данном месте и времени.

2.15. pathway — процесс, с помощью которого организм оказался в данном месте в данное время.

2.16. georeferenceVerificationStatus — категориальное описание степени, в которой географическая привязка была проверена для представления наилучшего возможного пространственного описания местоположения наблюдения.

2.17. occurrenceStatus — заявление о наличии или отсутствии таксона в данном месте.

2.18. preparations — список (объединенный и раздельный) препаратов и методов консервации экземпляра.

2.19. disposition — текущее состояние экземпляра по отношению к коллекции, идентифицированной в collectionCode или collectionID.

2.20. associatedMedia — список (объединенный и раздельный) идентификаторов (публикация, глобальный уникальный идентификатор, URI) носителей, связанных с наблюдением.

2.21. associatedOccurrences — список (объединенный и раздельный) идентификаторов других записей наблюдения и их связей с этим наблюдением.

2.22. associatedReferences — список (объединенный и раздельный) идентификаторов (публикация, библиографическая ссылка, глобальный уникальный идентификатор, URI) литературы, связанной с наблюдением.

2.23. associatedSequences — список (объединенный и раздельный) идентификаторов (публикация, глобальный уникальный идентификатор, URI) информации о генетической последовательности, связанной с наблюдением.

2.24. associatedTaxa — список (объединенный и раздельный) идентификаторов или названий таксонов и ассоциаций данного наблюдения к каждому из них.

2.25. otherCatalogNumbers — список (объединенный и раздельный) предыдущих или альтернативных полных номеров по каталогу или других используемых человеком идентификаторов для одного и того же наблюдения, как в текущем, так и в любом другом наборе данных или коллекции.

2.26. occurrenceRemarks — комментарии или примечания о наблюдении.

3. Organism

Категория описывает конкретный организм или определенную группу организмов, которая считается таксономически однородной.

3.1. organismID — идентификатор экземпляра организма (в отличие от конкретной цифровой записи организма), может быть глобально уникальным идентификатором или идентификатором, специфичным для набора данных.

3.2. organismName — текстовое имя или метка, назначенная экземпляру.

3.3. organismScope — описание типа экземпляра, может использоваться для указания того, представляет ли экземпляр дискретный организм или он представляет конкретный тип агрегации.

3.4. associatedOrganisms — список (объединенный и раздельный) идентификаторов других организмов и ассоциаций этого организма к каждому из них.

3.5. previousIdentifications — список (объединенный и раздельный) предыдущих названий организма.

3.6. organismRemarks — комментарии или заметки об экземпляре.

4. MaterialSample

Категория отражает физический результат сбора пробы. В биологических коллекциях образец материала обычно собирают и либо сохраняют, либо подвергают деструктивной обработке.

4.1. materialSampleID — идентификатор MaterialSample (в отличие от определенной цифровой записи образца материала), при отсутствии постоянного глобального уникального идентификатора создайте его на основе комбинации идентификаторов в записи, что в наибольшей степени сделает идентификатор matureSampleID глобально уникальным.

5. Event

Категория описывает действие, которое происходит в каком-либо месте в течение некоторого времени.

5.1. eventID — идентификатор для набора информации, связанной с событием (что-то, что происходит в данном месте и данное время), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.

5.2. parentEventID — идентификатор для более широкого события, которое группирует это и потенциально другие события.

5.3. fieldNumber — идентификатор события в поле, часто служит связующим звеном между полевыми заметками полей и событием.

5.4. eventDate — дата-время или интервал, в течение которого произошло событие, для наблюдений это дата-время записи события, не подходит для времени в геологическом контексте.

- 5.5. eventTime** — время или интервал, в течение которого произошло событие.
- 5.6. startDayOfYear** — самый ранний целый день года, в который произошло событие (1 за 1 января, 365 за 31 декабря, за исключением високосного года, в этом случае это 366).
- 5.7. endDayOfYear** — последний целый день года, в который произошло событие (1 за 1 января, 365 за 31 декабря, за исключением високосного года, в этом случае это 366).
- 5.8. year** — четырехзначный год, в котором произошло событие, согласно календарю нашей эры.
- 5.9. month** — целое число месяцев, в которых произошло событие.
- 5.10. day** — целое число дня месяца, в которое произошло событие.
- 5.11. verbatimEventDate** — дословное исходное представление информации о дате и времени события.
- 5.12. habitat** — категория или описание среды обитания, в которой произошло событие.
- 5.13. samplingProtocol** — названия, ссылки или описания методов или протоколов, используемых во время события.
- 5.14. sampleSizeValue** — числовое значение для измерения объема (продолжительность, длина, площадь или объем) выборки в событии выборки.
- 5.15. sampleSizeUnit** — единица измерения объема (продолжительность, длина, площадь или объем) выборки в событии выборки.
- 5.16. samplingEffort** — объем усилий, затраченных во время события (см. раздел 5).
- 5.17. fieldNotes** — один из а) индикатор существования, б) ссылка на (публикацию, URI) или в) текст записок, сделанных в поле о событии.
- 5.18. eventRemarks** — комментарии или заметки о событии.

6. Location

Данная категория терминов охватывает территориальную характеристику.

- 6.1. locationID** — идентификатор набора информации о местоположении (данные, связанные с dcterms: Location), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.
- 6.2. higherGeographyID** — идентификатор географического региона, в котором находится место проведения наблюдения.
- 6.3. higherGeography** — перечень (объединенный и раздельный) географических названий, менее конкретных, чем информация, собранная в терминах населенного пункта.
- 6.4. continent** — название континента, на котором находится место проведения наблюдения.
- 6.5. waterBody** — название водного объекта, в котором находится место проведения наблюдения.
- 6.6. islandGroup** — название группы островов, в которой находится место проведения наблюдения.
- 6.7. island** — название острова, на котором находится место проведения наблюдения или рядом с ним.
- 6.8. country** — название страны или основной административной единицы, в которой находится место проведения наблюдения.
- 6.9. countryCode** — стандартный код страны, в которой находится место проведения наблюдения.
- 6.10. stateProvince** — название следующего в сторону уменьшения уровня административного региона, чем страна (штат, провинция, кантон, департамент, регион и т. д.), в котором находится место проведения наблюдения.
- 6.11. county** — полное, не сокращенное название следующего в сторону уменьшения уровня административного района, чем stateProvince (графство, шир, департамент и т. д.), в котором находится место проведения наблюдения.
- 6.12. municipality** — полное, не сокращенное название следующего в сторону уменьшения административного района, чем уезд (город, муниципалитет и т. д.), в котором находится место проведе-

ния наблюдения, не используйте этот термин для ближайшего обозначенного места, не содержащего фактического местоположения.

6.13. locality — конкретное описание места проведения наблюдения.

6.14. verbatimLocality — исходное текстовое описание места проведения наблюдения.

6.15. minimumElevationInMeters — нижняя граница возвышенности (высота, обычно над уровнем моря), в метрах.

6.16. maximumElevationInMeters — верхний предел возвышенности (высота, обычно над уровнем моря), в метрах.

6.17. verbatimElevation — первоначальное описание возвышения (высоты, обычно над уровнем моря) Локации.

6.18. verticalDatum — вертикальный опорный элемент, используемый в качестве привязки, на которой основаны значения в терминах отметки.

6.19. minimumDepthInMeters — наименьшая глубина диапазона глубин ниже локальной поверхности, в метрах.

6.20. maximumDepthInMeters — наибольшая глубина диапазона глубин ниже локальной поверхности, в метрах.

6.21. verbatimDepth — исходное описание глубины под локальной поверхностью.

6.22. minimumDistanceAboveSurfaceInMeters — наименьшее расстояние в диапазоне расстояний от опорной поверхности в вертикальном направлении, в метрах, используйте положительные значения для местоположений над поверхностью, отрицательные значения для местоположений ниже ее, если измерения глубины даны, опорная поверхность является местоположением, заданным глубиной, в противном случае опорная поверхность является расположением, заданным отметкой.

6.23. maximumDistanceAboveSurfaceInMeters — наибольшее расстояние в диапазоне расстояний от опорной поверхности в вертикальном направлении, в метрах, используйте положительные значения для местоположений над поверхностью, отрицательные значения для местоположений ниже ее, если измерения глубины даны, опорная поверхность является местоположением, заданным глубиной, в противном случае опорная поверхность является расположением, заданным отметкой.

6.24. locationAccordingTo — сведения об источнике информации о местоположении, может быть изданием (справочником), учреждением или командой специалистов.

6.25. locationRemarks — комментарии или заметки о местоположении.

6.26. decimalLatitude — географическая широта (в десятичных градусах, с использованием системы пространственных привязок, заданной в `geodeticDatum`) географического центра местоположения, положительные значения — к северу от Экватора, отрицательные — к югу от него, значения находятся в диапазоне от -90 до 90 включительно.

6.27. decimalLongitude — географическая долгота (в десятичных градусах с использованием системы пространственных привязок, заданной в `geodeticDatum`) географического центра местоположения, положительные значения — к востоку от Гринвичского меридиана, отрицательные — к западу от него, значения составляют от -180 до 180 включительно.

6.28. geodeticDatum — эллипсоидная, геодезическая опорная или пространственная опорная система (SRS), на которой основаны географические координаты, заданные в `decimalLatitude` и `decimalLongitude`.

6.29. coordinateUncertaintyInMeters — расстояние по горизонтали (в метрах) от заданных `decimalLatitude` и `decimalLongitude`, описывающих наименьший круг, содержащий все местоположение, оставьте значение пустым, если неопределенность неизвестна, не может быть оценена или неприменима (поскольку отсутствуют координаты), нулевое значение не является допустимым для этого термина.

6.30. coordinatePrecision — десятичное представление точности координат, заданных в десятичной широте и десятичной долготе.

6.31. pointRadiusSpatialFit — отношение площади радиуса точки (`decimalLatitude`, `decimalLongitude`, `coordinateCoordinateInMeters`) к площади истинного (исходного или наиболее специфического) пространственного представления местоположения, допустимые значения равны 0, больше или равны 1 или не определены. Значение 1 является точным совпадением или 100% перекрытием, если заданный радиус точки не полностью содержит исходное представление, следует использовать значение 0, если исходное представление является точкой без неопределенности, а заданная географическая привязка не является той же точкой (без неопределенности), не определено (и должно быть оставлено пустым), если и исходная, и заданная географические привязки являются одной и той же точкой, в качестве точки `RadRadFit` используется 1.

6.32. verbatimCoordinates — дословные исходные пространственные координаты местоположения. Эллипсоид координат, геодетический опорный элемент или полная пространственная опорная система (SRS) для этих координат должны храниться в дословном SRS, а система координат — в дословном `CoordinateSystem`.

6.33. verbatimLatitude — Дословная исходная широта местоположения, эллипсоид координат, геодетический опорный элемент или полная пространственная опорная система (SRS) для этих координат должны храниться в дословном SRS, а система координат — в дословном `CoordinateSystem`.

6.34. verbatimLongitude — дословная исходная долгота местоположения, эллипсоид координат, геодетический опорный элемент или полная пространственная опорная система (SRS) для этих координат должны храниться в дословном SRS, а система координат — в дословном `CoordinateSystem`.

6.35. verbatimCoordinateSystem — формат координат для дословной широты и дословной долготы или дословных координат местоположения.

6.36. verbatimSRS — эллипсоид, геодезический опорный элемент или пространственная система отсчета (SRS), на которой основаны координаты, заданные в дословной широте и дословной долготы, или дословные координаты.

6.37. footprintWKT — Well-Known Text (WKT) — представление фигуры (компоновочный образ, геометрия), определяющее местоположение, расположение может иметь как представление радиуса точки (см. `decimalLatitude`), так и представление компоновочного образа, и они могут отличаться друг от друга.

6.38. footprintSRS — эллипсоид, геодезический опорный элемент или пространственная система координат (SRS), на которой основана геометрия, заданная в `footprintWKT`.

6.39. footprintSpatialFit — отношение площади компоновочного образа (`footprintWKT`) к площади истинного (исходного или наиболее специфического) пространственного представления местоположения, допустимые значения равны 0, больше или равны 1 или не определены, значение 1 является точным совпадением или 100% перекрытием, если данный компоновочный образ не полностью содержит исходное представление, следует использовать значение 0, `footprintFit` не определен (и должен быть оставлен пустым), если исходное представление является точкой без неопределенности, а данная географическая привязка не является этой точкой (без неопределенности), если исходная и заданная географические привязки совпадают, значение `footprintFit` равно 1.

6.40. georeferencedBy — список (объединенный и раздельный) имен людей, групп или организаций, которые определили географическую привязку (пространственное представление) для местоположения.

6.41. georeferencedDate — дата, на которую была сделана географическая ссылка на местоположение.

6.42. georeferenceProtocol — описание или ссылка на методы, используемые для определения пространственного пространства, координат и неопределенностей.

6.43. georeferenceSources — список (объединенный и отдельный) карт, справочников или других ресурсов, используемых для географической привязки местоположения, описанных достаточно конкретно, чтобы позволить любому человеку в будущем использовать те же ресурсы.

6.44. georeferenceRemarks — примечания или комментарии в отношении определения пространственного описания, поясняющие допущения, сделанные в дополнение или в противовес тем, которые формализованы в методе, упомянутом в Протоколе по географическому пространству.

7. GeologicalContext

Данная категория охватывает геологическую информацию, например стратиграфию, которая квалифицирует регион или место.

7.1. geologicalContextID — идентификатор для набора информации связанного с Geological Context (местоположение в геологическом контексте, таком как стратиграфия), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.

7.2. earliestEonOrLowestEonothem — полное название самой ранней из возможных геохронологических эонов или самой низкой хроно-стратиграфической эонотемы или неофициальное название («докембрий»), относящееся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.3. latestEonOrHighestEonothem — полное название новейшего возможного геохронологического эона или высшей хроно-стратиграфической эонотемы или неофициального названия («докембрий»), относящегося к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.4. earliestEraOrLowestErathem — полное название самой ранней возможной геохронологической эры или самой низкой хроностратиграфической эратемы, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.5. latestEraOrHighestErathem — полное название последней возможной геохронологической эры или наивысшей хроностратиграфической эратемы, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.6. earliestPeriodOrLowestSystem — полное название самого раннего возможного геохронологического периода или самой низкой хроностратиграфической системы, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный элемент.

7.7. latestPeriodOrHighestSystem — полное название последнего возможного геохронологического периода или наивысшей хроностратиграфической системы, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.8. earliestEpochOrLowestSeries — полное название самой ранней возможной геохронологической эпохи или самой низкой хроностратиграфической серии, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.9. latestEpochOrHighestSeries — полное название последней возможной геохронологической эпохи или наивысшей хроностратиграфической серии, относящейся к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.10. earliestAgeOrLowestStage — полное название самого раннего возможного геохронологического возраста или самого низкого хроностратиграфического этапа, относящегося к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.11. latestAgeOrHighestStage — полное название последнего возможного геохронологического возраста или наивысшего хроностратиграфического этапа, относящегося к стратиграфическому горизонту, с которого был собран каталогизированный предмет.

7.12. lowestBiostratigraphicZone — полное название минимально возможной геологической био-стратиграфической зоны стратиграфического горизонта, с которого был собран каталогизированный предмет.

7.13. highestBiostratigraphicZone — полное название максимально возможной геологической био-стратиграфической зоны стратиграфического горизонта, с которого был собран каталогизированный предмет.

7.14. lithostratigraphicTerms — сочетание всех лито-стратиграфических названий для породы, из которой был собран каталогизированный предмет.

7.15. group — полное название литостратиграфической группы, из которой был собран каталогизированный элемент.

7.16. formation — полное название литостратиграфической формации, из которой был собран каталогизированный предмет.

7.17. member — полное имя литостратиграфического члена, из которого был собран каталогизированный элемент.

7.18. bed — полное название литостратиграфического ложа, с которого был собран каталогизированный предмет.

8. Identification

Данная категория охватывает таксономические определения (например, отнесение к определенному таксону).

8.1. identificationID — идентификатор идентификации (совокупность информации, связанной с присвоением научного имени), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.

8.2. verbatimIdentification — строка, представляющая таксономическую идентификацию, как она появилась в исходной записи.

8.3. identificationQualifier — краткая фраза или стандартный термин («cf.», «aff.») для выражения сомнений определителя относительно Идентификации.

8.4. typeStatus — список (объединенный и раздельный) номенклатурных типов (статус типа, типизированное научное название, публикация).

8.5. identifiedBy — список (объединенный и раздельный) имен людей, групп или организаций, установивших таксономическую принадлежность субъекта.

8.6. identifiedByID — список (объединенный и раздельный) глобального уникального идентификатора для специалиста, людей, групп или организаций, ответственных за присвоение таксона субъекту.

8.7. dateIdentified — дата, на которую субъект был определен как представляющий таксон.

8.8. identificationReferences — список (объединенный и раздельный) ссылок (публикация, глобальный уникальный идентификатор, URI), используемых в идентификации.

8.9. identificationVerificationStatus — категорический индикатор степени, до которой была проверена корректная таксономическая идентификация, чтобы быть правильной.

8.10. identificationRemarks — комментарии или примечания по идентификации.

9. Taxon

Категория охватывает группы организмов (sensu http://purl.obolibrary.org/obo/OBI_0100026), рассматриваемых систематиками в качестве однородной единицы.

9.1. taxonID — идентификатор набора информации о таксоне (данные, связанные с классом Taxon), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.

- 9.2. scientificNameID** — идентификатор номенклатурных (не таксономических) деталей научного названия.
- 9.3. acceptedNameUsageID** — идентификатор использования имени, (документированное значение названия согласно источнику) действующего в настоящее время (зоологического) или принятого (ботанического) таксона.
- 9.4. parentNameUsageID** — идентификатор для использования имени (зарегистрированное значение имени согласно источнику) ближайшего таксона родителя более высокого ранга (в классификации) самого определенного элемента scientificName.
- 9.5. originalNameUsageID** — идентификатор для использования имени (зарегистрированное значение имени согласно источнику), в котором предельный элемент scientificName был первоначально установлен по правилам связанного pomenclaturalCode.
- 9.6. nameAccordingToID** — идентификатор источника, в котором определена или подразумевается конкретная концепция таксона (см. nameAccordingTo).
- 9.7. namePublishedInID** — идентификатор для публикации, в которой scientificName был первоначально установлен по правилам связанного pomenclaturalCode.
- 9.8. taxonConceptID** — идентификатор таксономического понятия, к которому относится запись — не для номенклатурных деталей таксона.
- 9.9. scientificName** — полное научное название с авторством и информацией о дате, если известно, при формировании части идентификации это должно быть имя в таксономическом ранге самого низкого уровня, которое может быть определено, этот термин не должен содержать идентификационных квалификаций, которые вместо этого должны быть предоставлены в термине ClausingQualifier.
- 9.10. acceptedNameUsage** — полное название с указанием авторства и даты, если известно, действующего в настоящее время (зоологического) или принятого (ботанического) таксона.
- 9.11. parentNameUsage** — полное имя с авторством и информацией о дате, если известно, ближайшего таксона родителя более высокого ранга (в классификации) самого определенного элемента scientificName.
- 9.12. originalNameUsage** — название таксона с информацией об авторстве и дате, если оно известно, как оно первоначально появилось при первом установлении по правилам ассоциированного номенклатурного кодекса. basionym (ботаника) или basonym (бактериология) scientificName или старшего/более раннего омонима для замененных имен.
- 9.13. nameAccordingTo** — Ссылка на источник, в котором определено или подразумевается конкретное понятие таксона — традиционно обозначаемое латинским «sensu» или «sec» (от secundum, означающее «по»), для таксонов, которые являются результатом идентификации, следует дать ссылку на ключи, монографии, экспертов и другие источники.
- 9.14. namePublishedIn** — ссылка для публикации, в которой scientificName был первоначально установлен по правилам связанного pomenclaturalCode.
- 9.15. namePublishedInYear** — год с четырьмя цифрами, в котором было опубликовано scientificName.
- 9.16. higherClassification** — список (объединенный и отдельный) названий таксонов, заканчивающихся в ранге, непосредственно превосходящем таксон, упоминаемый в записи таксона.
- 9.17. kingdom** — полное научное название царства, в котором классифицируется таксон.
- 9.18. phylum** — полное научное название типа или раздела, в котором классифицируется таксон.
- 9.19. class** — полное научное название класса, в котором классифицируется таксон.
- 9.20. order** — полное научное название отряда, в котором классифицируется таксон.
- 9.21. family** — полное научное название семейства, в котором классифицируется таксон.
- 9.22. subfamily** — полное научное название подсемейства, в котором классифицируется таксон.
- 9.23. genus** — полное научное название рода, в котором классифицируется таксон.

9.24. genericName — часть рода `scientificName` без авторства.

9.25. subgenus — полное научное название подрода, в котором классифицируется таксон, значения должны включать род, чтобы избежать путаницы гомонимов.

9.26. infragenericEpithet — промежуточная часть биномиального названия в рангах выше вида, но ниже рода.

9.27. specificEpithet — название первого эпитета или эпитета разновидностей `scientificName`.

9.28. infraspecificEpithet — название самого низкого или предельного внутривидового эпитета `scientificName`, исключая любое обозначение разряда.

9.29. cultivarEpithet — часть названия культивара, культиварной группы или грекса, которая следует за научным названием.

9.30. taxonRank — таксономический ранг собственного имени в `scientificName`.

9.31. verbatimTaxonRank — таксономический разряд собственного имени в `scientificName`, в оригинале.

9.32. scientificNameAuthorship — информация об авторстве для `scientificName` отформатирована согласно конвенциям применимого `nomenclaturalCode`.

9.33. vernacularName — распространенное или народное имя.

9.34. nomenclaturalCode — номенклатурный код (или коды в случае имени `ambiregna`), под которым построен `scientificName`.

9.35. taxonomicStatus — статус использования `scientificName` как этикетка для таксона, требует таксономического мнения для определения сферы применения таксона, правила приоритета затем используются для определения таксономического статуса номенклатуры, содержащейся в этой области, в сочетании с мнением экспертов, он должен быть связан с определенной таксономической ссылкой, определяющей понятие.

9.36. nomenclaturalStatus — статус, связанный с первоначальной публикацией названия и его соотношением с соответствующими правилами номенклатуры. Он основан по существу на алгоритме в соответствии с бизнес-правилами кода, не требует таксономического мнения.

9.37. taxonRemarks — комментарии или заметки о таксоне или названии.

10. MeasurementOrFact

Данная категория охватывает измерение или факт о `rdfs:Resource` (<http://www.w3.org/2000/01/rdf-schema#Resource>).

10.1. measurementID — идентификатор для `MeasurementOrFact` (информация, имеющая отношение к измерениям, фактам, особенностям или утверждениям), может быть глобальным уникальным идентификатором или идентификатором, специфичным для набора данных.

10.2. measurementType — характер измерения, факт, характеристика или утверждение.

10.3. measurementValue — значение измерения, факта, характеристики или утверждения.

10.4. measurementAccuracy — описание потенциальной ошибки, связанной с параметром `measurementValue`.

10.5. measurementUnit — единицы измерения, связанные с `measurementValue`.

10.6. measurementDeterminedBy — список (объединенный и отдельный) имен людей, групп или организаций, которые определили значение `MeasurementOrFact`.

10.7. measurementDeterminedDate — дата, в которую было сделано `MeasurementOrFact`.

10.8. measurementMethod — описание или ссылка (публикация, URI) на метод или протокол, используемый для определения измерения, факта, характеристики или утверждения.

10.9. measurementRemarks — комментарии или примечания, сопровождающие `MeasurementOrFact`.

5. Типы наборов данных, публикуемых в GBIF, и их структура

Наборы данных, публикуемые в GBIF, могут быть разделены на 4 типа, каждый из которых имеет свои особенности.

5.1. Metadata only

Данный тип наборов данных по сути представляет собой описание датасета. Он используется в случае необходимости описания какого-либо массива данных, коллекции и т. д., без предоставления фактических данных. Такой подход позволяет обозначить для широкого круга заинтересованных сам факт существования такого массива, в том числе сориентировать на его перспективную оцифровку и вовлечение в научный оборот через систему GBIF. Как и любой другой набор данных, датасеты данного типа также сопровождаются уникальным DOI.

Также следует обратить внимание, что рассмотренные далее три типа наборов данных в обязательном порядке включают в себя описательную часть в виде метаданных (см. раздел 8).

Список существующих датасетов данного типа можно посмотреть по ссылке <https://www.gbif.org/dataset/search?type=METADATA>. На момент подготовки руководства в GBIF зарегистрировано 364 подобного рода наборов.

5.2. Checklist data

Данный тип наборов данных представляет собой перечни таксонов, объединенных по какому-либо признаку, и как правило представляют собой простые списки, иногда с добавлением некоторой специфической информации, например, местных названий таксонов, синонимии, библиографических ссылок и т. д. Примером подобного рода наборов данных могут быть списки какого-либо региона, охраняемой территории и т. д.

В настоящее время в GBIF содержится более 40,5 тыс наборов данных этого типа, со списком которых можно ознакомиться по ссылке <https://www.gbif.org/dataset/search?type=CHECKLIST>.

Обязательными полями в шаблоне набора данных этого типа должны быть:

- taxonID — уникальный для формируемого набора данных код таксона, который сохраняется во всех последующих его версиях;
- scientificName — полное научное название таксона, включая авторство и год описания (если эта информация приемлема для таксона), в зависимости от типа предоставляемой информации может включать любой таксономический ранг, хотя в классическом варианте речь идет о видовом уровне;
- taxonRank — указывается, о каком таксономическом ранге идет речь для каждой записи.

В качестве настоятельно рекомендуемых предлагается добавлять:

- kingdom — и по возможности таксоны более высокого ранга;
- parentNameUsageID — уникальный идентификатор родительского таксона, используемый в текущем наборе данных;
- acceptedNameUsageID — уникальный для формируемого набора данных идентификатор ассоциированных названий, в том числе синонимов.

Также, как и любом другом типе наборов данных, рекомендуется при их составлении использовать как можно больше подходящих терминов Darwin Core, которые максимально полно представят информацию в создаваемом датасете.

Базовый шаблон для checklist набора данных (рисунок 23), а также шаблон с примерами (рисунки 24–27) доступны по ссылке <https://ipt.gbif.org/manual/en/ipt/2.5/checklist-data#templates>.

	A	B	C	D	E	F	G
1	taxonID	parentNameUsageID	acceptedNameUsageID	scientificName	kingdom	taxonRank	
2							
3							

Рисунок 23

A	B	C	D	E	F
taxonID	parentNameUsagID	parentNameUsage	acceptedNameUsagID	acceptedNameUsag	scientificName
1	73		73	Equisetopsida C. Aghard	Equisetopsida C. Aghard
2			26	Equisetidae Warming	Equisetidae Warming
3	26	26	25	Equisetales de Candolle ex Berchtold & J. Presl	Equisetales de Candolle ex Berchtold & J. Presl
4	25	25	128	Equisetaceae Michaux ex de Candolle	Equisetaceae Michaux ex de Candolle
5	128	128	1142	Equisetum Linnaeus	Equisetum Linnaeus
6	1142	1142	2004	Equisetum subg. Equisetum	Equisetum subg. Equisetum
7	2004	2004	5467	Equisetum fluviatile Linnaeus	Equisetum fluviatile Linnaeus
8	5467	2004	5466	Equisetum arvense Linnaeus	Equisetum arvense Linnaeus
9	5466	2004	5472	Equisetum pratense Ehrhart	Equisetum pratense Ehrhart
10	5472	2004	5471	Equisetum palustre Linnaeus	Equisetum palustre Linnaeus
11	5471	2004	5474	Equisetum sylvaticum Linnaeus	Equisetum sylvaticum Linnaeus
12	5474	2004	5482	Equisetum xilitorale Kühlewein ex Ruprecht	Equisetum xilitorale Kühlewein ex Ruprecht
13	5482	2004	5476	Equisetum telmateia Ehrhart	Equisetum telmateia Ehrhart
14	5476	5476	15836	Equisetum telmateia subsp. braunii (J. Milde) Hauke	Equisetum telmateia subsp. braunii (J. Milde) Hauke
15	15836	2004	5481	Equisetum xfont-queri Rothmaler	Equisetum xfont-queri Rothmaler
16	5481	1142	2005	Equisetum subg. Hippochaete (J. Milde) Baker	Equisetum subg. Hippochaete (J. Milde) Baker
17	2005	2005	5473	Equisetum scirpoides Michaux	Equisetum scirpoides Michaux
18	5473	2005	5484	Equisetum xnelsonii (A.A. Eaton) J.H. Schaffner	Equisetum xnelsonii (A.A. Eaton) J.H. Schaffner
19	5484	2005	5478	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	Equisetum variegatum Schleicher ex F. Weber & D. Mohr
20	5478	5478	5477	Equisetum variegatum subsp. alaskanum (A.A. Eaton) Hauke	Equisetum variegatum subsp. alaskanum (A.A. Eaton) Hauke
21	5477	5478	5479	Equisetum variegatum Schleicher ex F. Weber & D. Mohr	Equisetum variegatum Schleicher ex F. Weber & D. Mohr
22	5479	2005	5470	Equisetum laevigatum A. Braun	Equisetum laevigatum A. Braun
23	5470	2005	5480	Equisetum xferriisii Clute	Equisetum xferriisii Clute
24	5480	2005	5469	Equisetum hymale Linnaeus	Equisetum hymale Linnaeus
25	5469	5469	5468	Equisetum hymale subsp. affine (Engelmann) Calder & Roy L. Taylor	Equisetum hymale subsp. affine (Engelmann) Calder & Roy L. Taylor
26	5468	2005	5483	Equisetum xmackayi (Newman) Brichan	Equisetum xmackayi (Newman) Brichan
27	5483				
28					

Рисунок 24

A	G	H	I
nameAccordingToID	nameAccordingTo		kingdom
1	http://dx.doi.org/10.1111/j.1095-8339.2009.01001.0	Chase, M.W. & J.L. Reveal. 2009. A phylogenetic classification of land plants to accompany APG III. Botanical Journal of the Linnean Society 161 (2): 122-127.	Plantae
2	http://dx.doi.org/10.1111/j.1095-8339.2009.01001.0	Chase, M.W. & J.L. Reveal. 2009. A phylogenetic classification of land plants to accompany APG III. Botanical Journal of the Linnean Society 161 (2): 122-127.	Plantae
3	http://www.jstor.org/stable/25065646	Smith, A.R., K.M. Pryer, E. Schuettpelz, P. Korall, H. Schneider & P.G. Wolf. 2006. A classification for extant ferns. Taxon 55: 705-731.	Plantae
4	http://www.jstor.org/stable/25065646	Smith, A.R., K.M. Pryer, E. Schuettpelz, P. Korall, H. Schneider & P.G. Wolf. 2006. A classification for extant ferns. Taxon 55: 705-731.	Plantae
5	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
6	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
7	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
8	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
9	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
10	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
11	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
12	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
13	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
14	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
15	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
16	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
17	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
18	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
19	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
20	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
21	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
22	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
23	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
24	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
25	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
26	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
27	http://www.efloras.org/volume_page.aspx?volum	FNA Editorial Committee. 1993. Flora of North America north of Mexico. Volume 2: Pteridophytes and Gymnosperms. Oxford University Press, New York.	Plantae
28			

Рисунок 25 – продолжение таблицы на рисунке 24

A	J	K	L	M	N	O	P	Q	R	S	T
class	order	family	genus	subgenus	specificEpithet	infraspecificEpithet	taxonRank	scientificNameAuthorship	taxonomicStatus	modified	
1	Equisetopsida						class	C. Aghard	accepted	2011-02-21T12:19:0500	
2	Equisetopsida						subclass	Warming	accepted	2011-02-21T12:19:0500	
3	Equisetopsida	Equisetales					order	de Candolle ex Berchtold & J. Presl	accepted	2011-02-21T12:19:0500	
4	Equisetopsida	Equisetales	Equisetaceae				family	Michaux ex de Candolle	accepted	2011-02-21T12:19:0500	
5	Equisetopsida	Equisetales	Equisetaceae	Equisetum			genus	Linnaeus	accepted	2011-02-21T12:19:0500	
6	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum		subgenus		accepted	2011-02-21T12:19:0500	
7	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	fluviatile	species	Linnaeus	accepted	2011-02-21T12:20:0500	
8	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	arvense	species	Linnaeus	accepted	2011-02-21T12:20:0500	
9	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	pratense	species	Ehrhart	accepted	2011-02-21T12:20:0500	
10	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	palustre	species	Linnaeus	accepted	2011-02-21T12:20:0500	
11	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	sylvaticum	species	Linnaeus	accepted	2011-02-21T12:20:0500	
12	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	xilitorale	subspecies	(A.A. Eaton) Huldén	accepted	2011-02-21T12:20:0500	
13	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	telmateia	species	Ehrhart	accepted	2011-08-16T10:36:0400	
14	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	telmateia	subspecies	(J. Milde) Hauke	accepted	2011-08-16T10:36:0400	
15	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Equisetum	font-queri	species	Rothmaler	accepted	2011-02-21T12:20:0500	
16	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete		subgenus	(J. Milde) Baker	accepted	2011-02-21T12:19:0500	
17	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete		species	Michaux	accepted	2011-02-21T12:20:0500	
18	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	scirpoides	species	Linnaeus	accepted	2011-08-23T16:55:0400	
19	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	xnelsonii	species	(A.A. Eaton) J.H. Schaffner	accepted	2011-08-16T10:36:0400	
20	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	variegatum	species	Schleicher ex F. Weber & D. Mohr	accepted	2011-02-21T12:20:0500	
21	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	variegatum	subspecies	(A.A. Eaton) Huldén	accepted	2011-02-21T12:20:0500	
22	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	variegatum	variegatum	Schleicher ex F. Weber & D. Mohr	accepted	2011-02-21T12:20:0500	
23	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	laevigatum	species	A. Braun	accepted	2011-02-21T12:20:0500	
24	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	xferriisii	species	Clute	accepted	2011-02-21T12:20:0500	
25	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	hymale	species	Linnaeus	accepted	2011-08-16T10:36:0400	
26	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	hymale	subspecies	(Engelmann) Calder & Roy L. Taylor	accepted	2011-02-21T12:20:0500	
27	Equisetopsida	Equisetales	Equisetaceae	Equisetum	Equisetum subg. Hippochaete	xmackayi	species	(Newman) Brichan	accepted	2011-08-23T16:55:0400	
28											

Рисунок 26 – продолжение таблицы на рисунке 25

U	V
1 bibliographicCitation	references
2 Brouillet et al. 2010+. Equisetopsida C. Aghard in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/73 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/73
3 Brouillet et al. 2010+. Equisetidae Warming in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/26 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/26
4 Brouillet et al. 2010+. Equisetales Candolle ex Berchtold & J. Presl in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/25 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/25
5 Brouillet et al. 2010+. Equisetaceae Michaux ex de Candolle in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/128 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/128
6 Brouillet et al. 2010+. Equisetum Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/1142 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/1142
7 Brouillet et al. 2010+. Equisetum subg. Equisetum in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/2004 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/2004
8 Brouillet et al. 2010+. Equisetum flavatile Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5467 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5467
9 Brouillet et al. 2010+. Equisetum arvense Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5466 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5466
10 Brouillet et al. 2010+. Equisetum pratense Ehrhart in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5472 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5472
11 Brouillet et al. 2010+. Equisetum palustre Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5471 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5471
12 Brouillet et al. 2010+. Equisetum sylvaticum Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5474 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5474
13 Brouillet et al. 2010+. Equisetum x-litorale Kihlewein ex Ruprecht in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5482 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5482
14 Brouillet et al. 2010+. Equisetum telmateia Ehrhart in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5476 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5476
15 Brouillet et al. 2010+. Equisetum telmateia subsp. braunii (J. Milde) Hauke in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/15836 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/15836
16 Brouillet et al. 2010+. Equisetum xfont-queri Rothmaler in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5481 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5481
17 Brouillet et al. 2010+. Equisetum subg. Hippochaete (J. Milde) Baker in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/2005 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/2005
18 Brouillet et al. 2010+. Equisetum scirpoides Michaux in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5473 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5473
19 Brouillet et al. 2010+. Equisetum xnelsoni (A.A. Eaton) J.H. Schaffner in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5484 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5484
20 Brouillet et al. 2010+. Equisetum variegatum Schleicher ex F. Weber & D. Mohr in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5477 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5477
21 Brouillet et al. 2010+. Equisetum variegatum subsp. alaskanum (A.A. Eaton) Hultén in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5477 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5477
22 Brouillet et al. 2010+. Equisetum variegatum Schleicher ex F. Weber & D. Mohr subsp. variegatum in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5477 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5477
23 Brouillet et al. 2010+. Equisetum laevigatum A. Braun in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5470 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5470
24 Brouillet et al. 2010+. Equisetum xferissii Clute in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5480 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5480
25 Brouillet et al. 2010+. Equisetum hemale Linnaeus in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5469 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5469
26 Brouillet et al. 2010+. Equisetum hemale subsp. affinis (Engelmann) Calder & Roy L. Taylor in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5468 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5468
27 Brouillet et al. 2010+. Equisetum xmackayi (Newman) Brichan in VASCAN, the Database of Vascular Plants of Canada. http://data.canadensys.net/vascan/taxon/5483 (consulted on 2015-09-24)	http://data.canadensys.net/vascan/taxon/5483

Рисунок 27 — продолжение таблицы на рисунке 26

5.3. Occurrence data

Данные о находках являются наиболее широко представленным, но не самым многочисленным типом наборов данных. На момент подготовки окончательной версии руководства в системе GBIF зарегистрировано более 2,13 млрд записей наблюдений. Под наблюдением может пониматься либо непосредственное наблюдение (observation) за биологическим объектом в природе, и факт об этом зафиксирован в виде записи в полевом дневнике, снимке фотолушкы, записи в диктофоне, на смартфоне и т. д. Также под наблюдением будет пониматься информация об экземпляре представленном в коллекции, гербарии и т. д. (specimen), для которого приведены помимо его таксономического статуса дополнительные данные, отражаемые в этикетке. В последующем, при подготовке наборов данных к публикации в GBIF, эти факторы будут учитываться отдельно (см. раздел 8).

Со списком датасетов данного типа, которых в настоящее время в GBIF зарегистрировано более 24,8 тыс., можно ознакомиться по ссылке <https://www.gbif.org/dataset/search?type=OCCURRENCE>.

Обязательными полями в шаблоне набора данных этого типа будут:

- occurrenceID — уникальный идентификатор наблюдения (непосредственно наблюдения либо коллекционного экземпляра), позволяющий отличать его во всех последующих версиях набора данных;
- basisOfRecord — тип отдельной записи (например, наблюдение, физический образец, окаменелость и т. д.);
- scientificName — полное научное название максимально допустимо низкого для конкретного наблюдения таксономического ранга;
- eventDate — дата наблюдения.

В качестве настоятельно рекомендуемых предлагается добавлять:

- taxonRank — таксономический ранг представленного в конкретной записи наблюдения научного названия таксона;
- kingdom — царство, к которому относится приводимый таксон, также по возможности рекомендуется использовать таксон еще более высокого ранга;
- decimalLatitude & decimalLongitude & geodeticDatum — географические широта и долгота в десятичных градусах, а также используемая система координат;

- **contryCode** — двухбуквенная стандартная аббревиатура для страны, в которой было сделано наблюдение (согласно стандарту ISO 3166);
- **individualCount/organismQuantity & organismQuantityType** — сопутствующая информация о количестве экземпляров в конкретно взятом наблюдении, процента покрытия растительностью, биомассы и т. п.

Как и в случае с предыдущим типом наборов данных в системе GBIF по ссылке <https://ipt.gbif.org/manual/en/ipt/2.5/occurrence-data#templates> доступны базовый шаблон (рисунок 28) и пример набора данных наблюдений (рисунки 29–35).

	A	B	C	D	E	F	G	H	I	J	K	L	M
1	occurrenceID	basisOfRecord	eventDate	kingdom	scientificName	taxonRank	decimalLatitude	decimalLongitude	geodeticDatum	countryCode	individualCount	organismQuantity	organismQuantityType
2													
3													
4													
5													
6													
7													

Рисунок 28

	A	B	C	D	E	F	G	H	I	J
1	occurrenceID	basisOfRecord	eventDate	endDayOfYear	year	month	day	verbatimEventDate	eventRemarks	scientificName
2	http://arctos.database.museum/guid/CUMV:Amph:11733?seid=2231302	PreservedSpecimen	1926-04		1926	4		0/4/1926	day of month unknown	Ambystoma maculatum
3	http://arctos.database.museum/guid/CUMV:Amph:4335?seid=2239851	PreservedSpecimen	1942-04-17		1942	4	17	17.04.1942		Desmognathus fuscus
4	http://arctos.database.museum/guid/CUMV:Amph:4337?seid=2239852	PreservedSpecimen	1942-04-17		1942	4	17	17.04.1942		Gyrinophilus porphyriticus
5	http://arctos.database.museum/guid/CUMV:Amph:4340?seid=2243745	PreservedSpecimen	1942-04-17		1942	4	17	17.04.1942		Eurycea bislineata bislineata
6	http://arctos.database.museum/guid/CUMV:Amph:4341?seid=2246587	PreservedSpecimen	1942-04-17		1942	4	17	17.04.1942		Plethodon cinereus
7	http://arctos.database.museum/guid/CUMV:Amph:16004?seid=2329885	PreservedSpecimen	1953-09-27		1953	9	27	27.сеп.53		Rana sylvatica
8	http://arctos.database.museum/guid/CUMV:Amph:11618?seid=2241805	PreservedSpecimen	1979-06-02/1979-06-07					02.06.1979		Eleutherodactylus eneidiae
9	http://arctos.database.museum/guid/CUMV:Amph:16023?seid=2660598	PreservedSpecimen	1981-06-01		1981	6	1	01.июн.81		Masticophis flagellum plicatus
10	http://arctos.database.museum/guid/CUMV:Amph:16113?seid=2674154	PreservedSpecimen	2011-06-23		174	2011	6	23	23.июн.11	Rana (Lithobates) clamitans
11										
12										

Рисунок 29

	K	L	M	N	O	P	Q	R	S	T	U	V	W	X
1	higherClassification	kingdom	phylum	class	order	family	genus	specificEpithet	infraspecificEpithet	taxonRank	identifiedBy	dateIdentified	nomenclaturalCode	decimalLatitude
2	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Caudata	Ambystomatid	Ambystoma	maculatum			species	unknown	1926-04	ICZN	42.4566
3	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Caudata	Plethodontida	Desmognathus	fuscus			species	unknown	1942-04-17	ICZN	42.4566
4	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Caudata	Plethodontida	Gyrinophilus	porphyriticus			species	unknown	1942-04-17	ICZN	42.4566
5	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Caudata	Plethodontida	Eurycea	bislineata	bislineata		subspecies	unknown	1942-04-17	ICZN	42.4566
6	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Caudata	Plethodontida	Plethodon	cinereus			species	unknown	1942-04-17	ICZN	42.4566
7	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Anura	Ranidae	Rana	sylvatica			species	J. N. Layne		ICZN	42.4566
8	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Anura	Leptodactylidae	Eleutherodactylus	eneidiae			species	unknown	1979-06-02	ICZN	18.302
9	Animalia; Chordata; Repti Animalia	Chordata	Reptilia	Squamata	Colubridae	Masticophis	flagellum	plicatus		subspecies	Sean Mchugh		ICZN	32.7038
10	Animalia; Chordata; Ampl Animalia	Chordata	Amphibia	Anura	Ranidae	Rana	(Lithobates)	clamitans		species	Kelly R. Zamudio		ICZN	43.362
11														

Рисунок 30 — продолжение таблицы на рисунке 29

	Y	Z	AA	AB	AC	AD	AE	AF	AG
1	decimalLongitude	geodeticDatum	coordinateUncertaintyInMeters	verbatimCoordinates	verbatimCoordinateSystem	georeferencedBy	georeferencedDate	georeferenceProtocol	georeferenceSources
2	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	unknown	2014-03-03	MaNIS georeferencing g:	NamedPlaceExtent: 1000
3	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	unknown	2014-03-03	MaNIS georeferencing g:	NamedPlaceExtent: 1000
4	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	unknown	2014-03-03	MaNIS georeferencing g:	NamedPlaceExtent: 1000
5	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	unknown	2014-03-03	MaNIS georeferencing g:	NamedPlaceExtent: 1000
6	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	unknown	2014-03-03	MaNIS georeferencing g:	NamedPlaceExtent: 1000
7	-76.45442	WGS84	500	42.4566/-76.45442	decimal degrees	Charles M. Dardia	2014-05-16	MaNIS georeferencing g:	NamedPlaceExtent: 1000
8	-65.793	WGS84		18.302/-65.793	decimal degrees	unknown	2014-03-03	not recorded	Georeference Source: To
9	-109.9191	WGS84		542 32.7038/-109.9191	decimal degrees	Sean Mchugh	2015-07-15	BioGeoMancer	Beogeomancer
10	-74.587	WGS84		250 43.362/-74.587	decimal degrees	Charles M. Dardia	2015-09-18	not recorded	GPS
11									

Рисунок 31 — продолжение таблицы на рисунке 30

	AH	AI	AJ	AK	AL	AM	AN	AO	AP
1	georeferenceVerificationStatus	higherGeography	continent	islandGroup	island	country	countryCode	stateProvince	county
2	unverified	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
3	unverified	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
4	unverified	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
5	unverified	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
6	unverified	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
7	checked by collector	North America, United States, New York, Tompkins County	North America			United States	US	New York	Tompkins County
8	unverified	West Indies, United States, Puerto Rico, Greater Antilles	West Indies	Greater Antilles		United States	US	Puerto Rico	
9	unverified	North America, United States, Arizona, Graham County	North America			United States	US	Arizona	Graham County
10	unverified	North America, United States, New York, Hamilton County	North America			United States	US	New York	Hamilton County
11									

Рисунок 32 — продолжение таблицы на рисунке 31

	AQ	AR	AS	AT	AU	AV	AW
1	locality	verbatimLocality	locationAccordingTo	type	modified	language	license
2	Ithaca, Forest Home, CU Rifle Range, Noipond near rifle range		unknown	PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
3	Ithaca, Forest Home, CU Rifle Range, NoIRifle Range, Fall Cr., Ithaca,		unknown	PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
4	Ithaca, Forest Home, CU Rifle Range, NoIRifle Range, Fall Cr., Ithaca,		unknown	PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
5	Ithaca, Forest Home, CU Rifle Range, NoIRifle Range, Fall Cr., Ithaca,		unknown	PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
6	Ithaca, Forest Home, CU Rifle Range, NoIRifle Range, Fall Cr., Ithaca,		unknown	PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
7	Ithaca, Forest Home, CU Rifle Range, NoIthaca, Cornell University Rifle Rang	Charles M. Dardia		PhysicalObject	2015-11-13 12:43:50.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
8	Mt. Britton Trail	Mt. Britton Trail	unknown	PhysicalObject	2015-10-28 10:33:21.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
9	Head Soldier Creek, Graham Mts., 9300 FHead Soldier Creek, Graham Mts., 9	Sean Mchugh		PhysicalObject	2015-07-16 12:39:26.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode
10	Piseco, 11 km SSW, Powley Road	Powley Rd. Lake Pleasant, NY Adliro	Charles M. Dardia	PhysicalObject	2015-09-18 10:40:34.0	en	http://creativecommons.org/publicdomain/zero/1.0/legalcode

Рисунок 33 – продолжение таблицы на рисунке 32

	AX	AY	AZ	BA	BB	BC	BD	BE
1	references	institutionID	institutionCode	collectionCode	catalogNumber	occurrenceRemarks	recordNumber	recordedBy
2	http://arctos.database.museum/guid/CUMV:Amph:11733	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	11733	partly hatched		Collector(s): John R. Greeley
3	http://arctos.database.museum/guid/CUMV:Amph:4335	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	4335			Collector(s): Robert M. Roecker
4	http://arctos.database.museum/guid/CUMV:Amph:4337	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	4337			Collector(s): Robert M. Roecker
5	http://arctos.database.museum/guid/CUMV:Amph:4340	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	4340			Collector(s): Robert M. Roecker
6	http://arctos.database.museum/guid/CUMV:Amph:4341	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	4341			Collector(s): Robert M. Roecker
7	http://arctos.database.museum/guid/CUMV:Amph:16004	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	16004	in mouth of <i>Thamnophis sirtalis</i>		Collector(s): J. N. Layne
8	http://arctos.database.museum/guid/CUMV:Amph:11618	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	11618		PR 1687	Collector(s): J. Wilson, R. Thomas
9	http://arctos.database.museum/guid/CUMV:Amph:16023	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	16023	C.A. Beuchat collector		Collector(s): unknown
10	http://arctos.database.museum/guid/CUMV:Amph:16113	http://grbio.org/cool/164g-wjcr	CUMV	Amphibian specimens	16113		KZ 3386	Collector(s): Guilherme Becker

Рисунок 34 – продолжение таблицы на рисунке 33

	BF	BG	BH	BI	BJ	BK	BL
1	organismID	individualCount	organismQuantity	organismQuantity	establishmentMeans	preparations	otherCatalogNumbers
2	http://arctos.database.museum/guid/CUMV:Amph:11733	1	1 individuals	wild caught		egg (formalin)	Trapline ID=JRG 1926APR; original identifier=CU Egg 106
3	http://arctos.database.museum/guid/CUMV:Amph:4335	1	1 individuals	wild caught		whole organism (unknown)	Trapline ID=RMR 1942APR17
4	http://arctos.database.museum/guid/CUMV:Amph:4337	1	1 individuals	wild caught		whole organism (unknown)	Trapline ID=RMR 1942APR17
5	http://arctos.database.museum/guid/CUMV:Amph:4340	1	1 individuals	wild caught		whole organism (unknown)	Trapline ID=RMR 1942APR17
6	http://arctos.database.museum/guid/CUMV:Amph:4341	1	1 individuals	wild caught		whole organism (unknown)	Trapline ID=RMR 1942APR17
7	http://arctos.database.museum/guid/CUMV:Amph:16004	1	1 individuals	wild caught		whole organism (Isopropanol)	
8	http://arctos.database.museum/guid/CUMV:Amph:11618	1	1 individuals	wild caught		whole organism (Isopropanol)	Trapline ID=JW 1979JUN02A; collector number=PR 1687
9	http://arctos.database.museum/guid/CUMV:Amph:16023	1	1 individuals	wild caught		whole organism (mummified)	
10	http://arctos.database.museum/guid/CUMV:Amph:16113	1	1 individuals	wild caught		whole organism (ethanol)	collector number=KZ 3386

Рисунок 35 – продолжение таблицы на рисунке 34

5.4. Sampling-event data

Этот относительно недавно введенный в систему GBIF набор данных позволяет оперировать еще более широкими объемами информации о биологическом разнообразии. Фактически в пределах данных датасетов идет объединение данных о находках и подробного описания маршрутов, экспериментальных площадок и т. д. Такого рода наборы данных позволяют более свободно оперировать количественными данными.

К моменту подготовки данного руководства в GBIF зарегистрировано 2154 Sampling-event data датасетов. С их списком можно ознакомиться по ссылке https://www.gbif.org/dataset/search?type=SAMPLING_EVENT.

Обязательными полями в шаблоне набора данных этого типа будут:

- eventID – уникальный идентификатор события, который позволяет связать конкретное наблюдение с конкретным событием отбора проб, позволяющий отличать его во всех последующих версиях набора данных;
- eventDate – дата или интервал дат, в течение которого произошло событие в соответствии со стандартом и временем ISO 8601;
- samplingProtocol – название, ссылка либо описание метода или протокола, применяемого при сборе данных (согласно рекомендуемым терминам Darwin Core).

В качестве настоятельно рекомендуемых предлагается добавлять:

- sampleSizeValue & sampleSizeUnit – числовое значение и соответствующая единица измерения, указывающая размер отдельной выборки в событии отбора проб;
- parentEventID – перекрестная ссылка на более обширное ассоциированное событие, например проект мониторинга, в рамках которого происходил отбор проб;

- **samplingEffort** — мера усилий, затраченных на отбор проб, например, количество наблюдателей, количество ловушек и т. д.;
- **locationID** — внутренняя или внешняя ссылка на набор данных, описывающая местоположение, где происходило событие отбора проб;
- **decimalLatitude & decimalLongitude & geodeticDatum** — географические широта и долгота в десятичных градусах, а также используемая система координат;
- **footprintWKT&footprintSPS** — Well-known text. текстовое описание представления векторной графики и писания системы координат;
- **contryCode** — двухбуквенная стандартная аббревиатура для страны, в которой было сделано наблюдение (согласно стандарту ISO 3166);
- **occurrenceStatus** — дополнительные сведения по отбору проб, например, факт отсутствия таксона и т. д.

Как и в случае с предыдущими типами наборов данных, в системе GBIF по ссылке <https://ipt.gbif.org/manual/en/ipt/2.5/sampling-event-data#templates> доступны базовый шаблон события отбора проб (рисунок 36) с ассоциированной таблицей наблюдений (рисунки 37, 38) и пример набора данных события отбора проб (рисунки 39, 40) и связанных с ними наблюдений (рисунки 41, 42).

A	B	C	D	E	F	G	H	I	J	K	L	M	N	
1	eventID	parentEventID	eventDate	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	locationID	footprintWKT	footprintSPS	decimalLatitude	decimalLongitude	geodeticDatum	countryCode
2														
3														
4														
5														
6														
7														
8														

Рисунок 36

A	B	C	D	E	F	G	H	I	J	
1	eventID	occurrenceID	basisOfRecord	eventDate	kingdom	scientificName	taxonRank	decimalLatitude	decimalLongitude	geodeticDatum
2										
3										
4										

Рисунок 37

K	L	M	N	O	
1	countryCode	individualCount	organismQuantity	organismQuantityType	occurrenceStatus
2					
3					
4					

Рисунок 38 — продолжение таблицы на рисунке 37

A	B	C	D	E	F	G	H	I	
1	eventID	samplingProtocol	samplingEffort	sampleSizeValue	sampleSizeUnit	eventDate	eventTime	startDayOfYear	eventRemarks
2	994-tr009-s00	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2012-10-11 09:28:02Z/10:16:02Z		284	No occurrences
3	3502-tr056-s6	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19 12:25:02Z/13:10:02Z		291	
4	3502-tr056-s9	Pollard walks	Average of 30 Minutes walk along transect	250 square metre		2015-10-19 12:25:02Z/13:10:02Z		291	
5									

Рисунок 39

J	K	L	M	N	O	P	Q	R	S	
1	country	countryCode	locality	locationID	decimalLatitude	decimalLongitude	geodeticDatum	coordinateUncertaintyInMeters	type	ownerInstitutionCode
2	Israel	IL	Sde boker rechesh halukim	tr009-s00	30.8539	34.7692	WGS84	25	Event	ILS
3	Israel	IL	Nahal Kovshim Beer Sheva	tr056-s6	31.29077	34.81112	WGS84	25	Event	ILS
4	Israel	IL	Nahal Kovshim Beer Sheva	tr056-s9	31.28945	34.81151	WGS84	25	Event	ILS
5										

Рисунок 40 — продолжение таблицы на рисунке 39

	A	B	C	D	E	F	G	H
1	eventID	occurrenceID	basisOfRecord	individualCount	organismQuantity	organismQuantityType	occurrenceStatus	scientificName
2	1382-tr009-s00	1382-tr009-s00-0	HumanObservation	0	0 individuals		absent	Lepidoptera
3	3502-tr056-s6	3502-tr056-s6-21114	HumanObservation	3	3 individuals		present	Azanus jesous
4	3502-tr056-s6	3502-tr056-s6-21126	HumanObservation	1	1 individuals		present	Melitaea trivia
5	3502-tr056-s6	3502-tr056-s6-21127	HumanObservation	3	3 individuals		present	Deudorix livia
6	3502-tr056-s6	3502-tr056-s6-21129	HumanObservation	1	1 individuals		present	Azanus ubaldus
7	3502-tr056-s6	3502-tr056-s6-21132	HumanObservation	1	1 individuals		present	Lycaena thersamon
8	3502-tr056-s9	3502-tr056-s9-21116	HumanObservation	1	1 individuals		present	Azanus jesous
9	3502-tr056-s9	3502-tr056-s9-21122	HumanObservation	1	1 individuals		present	Tarucus balkanica
10	3502-tr056-s9	3502-tr056-s9-21131	HumanObservation	1	1 individuals		present	Azanus ubaldus
11								

Рисунок 41

	I	J	K	L	M	N	O	P	Q	R
1	kingdom	phylum	class	order	family	infraspecificEpithet	taxonRank	recordedBy	type	ownerInstitutionCode
2	Animalia	Arthropoda	Insecta	Lepidoptera			order	Eviatar Fingold	Event	ILS
3	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
4	Animalia	Arthropoda	Insecta	Lepidoptera	Nymphalidae		species	Zvika Avni	Event	ILS
5	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
6	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
7	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
8	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
9	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
10	Animalia	Arthropoda	Insecta	Lepidoptera	Lycaenidae		species	Zvika Avni	Event	ILS
11										

Рисунок 42 – продолжение таблицы на рисунке 41

6. Качество данных

Как уже подчеркивалось ранее, в рамках данного руководства не рассматривается весь спектр вопросов, касающихся работы с данными как таковыми, однако, на понятии качества данных подоготовленного к публикации датасета мы остановимся немного подробнее.

Безусловно, качество получаемых и предоставляемых данных должно постоянно находиться в центре внимания специалистов. Однако, общеизвестно, что независимо от прилагаемых усилий, особенно при больших массивах данных, исчисляемых десятками и сотнями тысяч записей, наличие различного рода ошибок практически неизбежно. Это касается не только сферы информатики биоразнообразия, но и любых сфер деятельности, где приходится иметь дело с большими объемами данных.

Для минимизации ошибок, помощи в их поиске и корректировке существуют различные инструменты, начиная от относительно универсальных, например, Excel, заканчивая специализированными, ориентированными как на работу с массивами данных как таковыми, например OpenRefine, так и на проверку специфической информации, в том числе относящейся непосредственно к биологическому разнообразию. В частности, такой инструментарий существует и в системе GBIF.

Довольно распространенной практикой, в том числе при работе по подготовке наборов данных для публикации в GBIF, является использование OpenRefine.

Этот мощный инструмент используется для так называемой очистки данных. Под этим понятием понимают очистку массива данных от ошибок, как технических, так и смысловых. В Интернете существует огромное количество руководств по работе с этой системой, в которых описываются и особенности установки на различные операционные системы и нюансы эксплуатации, поэтому в рамках настоящего издания мы их не рассматриваем.

Программа позволяет работать с данными различных форматов, в том числе с форматом csv, рекомендованном для загрузки наборов данных через IPT, о чем будет более подробно сказано далее в разделе 8.

Также можно воспользоваться существующим в GBIF инструментом — GBIF Data Validator (рисунки 43), перейти на который можно по ссылке <https://www.gbif.org/tools/data-validator>.

Инструментарий позволяет провести анализ подготовленного набора данных, предложить рекомендации по его улучшению и очистке. Файл можно загрузить с компьютера либо виртуального источника.

В частности при анализе тестового набора данных было отмечено, что не все поля заполнены, показана доля ошибок (рисунки 44–47).

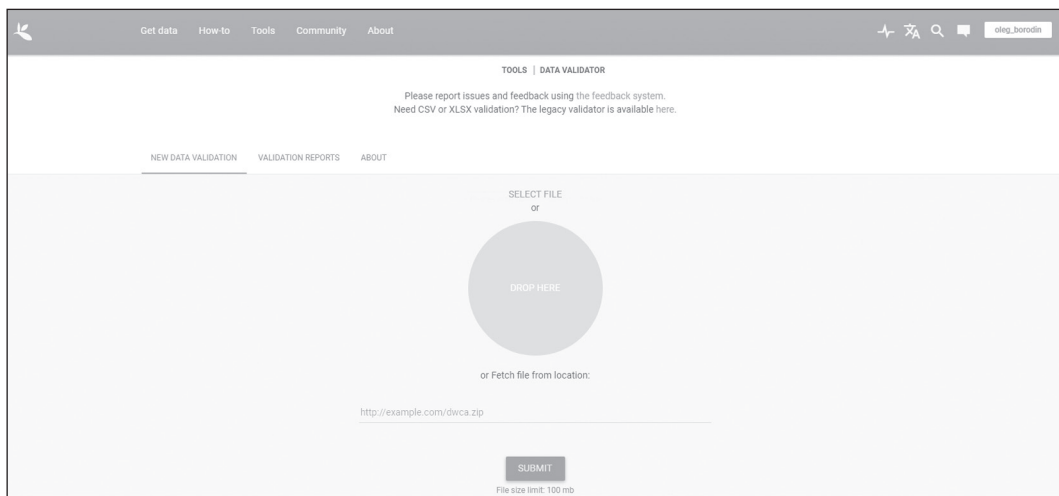


Рисунок 43

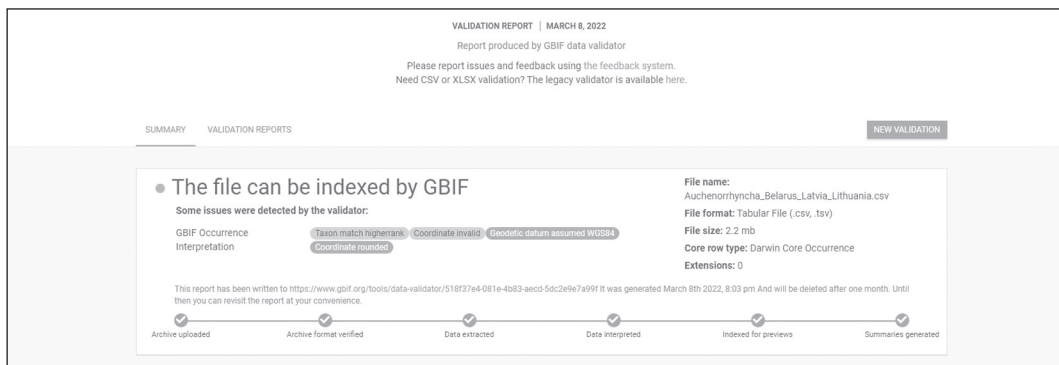


Рисунок 44

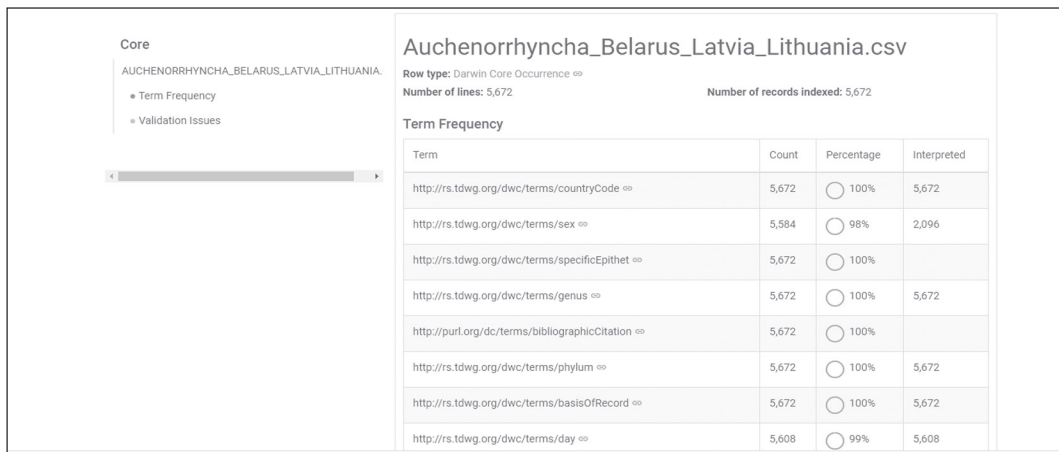


Рисунок 45

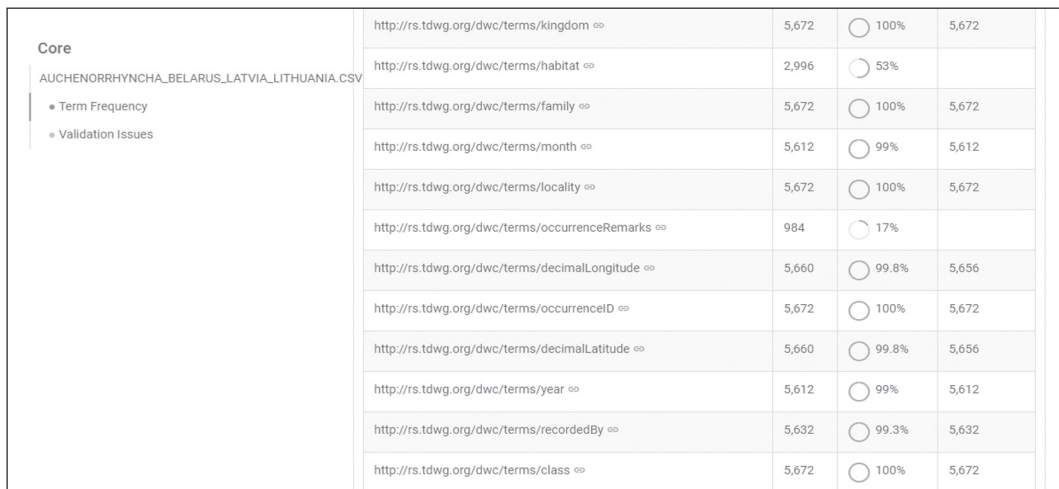


Рисунок 46

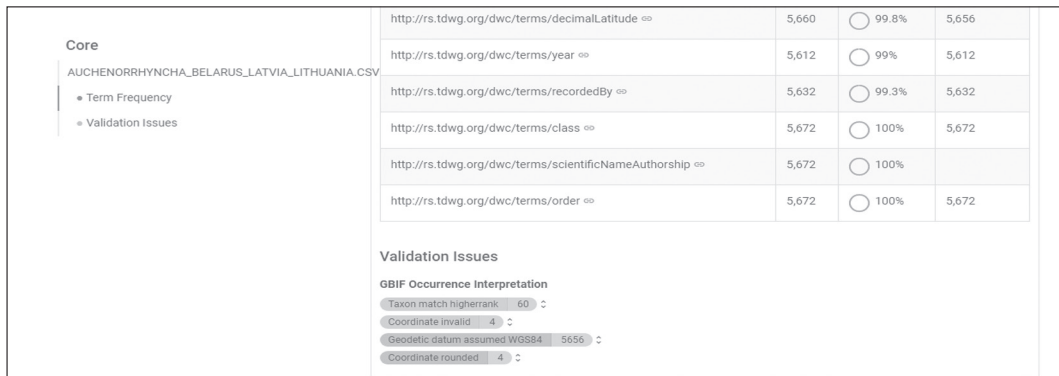


Рисунок 47

Как следствие, данные ошибки могут быть ценеправленно выявлены и устранены. В конечном итоге, по результатам проведенного анализа в отчете работы GBIF Data Validator содержится информация о пригодности набора данных для публикации (рисунок 48), ссылка на отчет присылается и на e-mail.

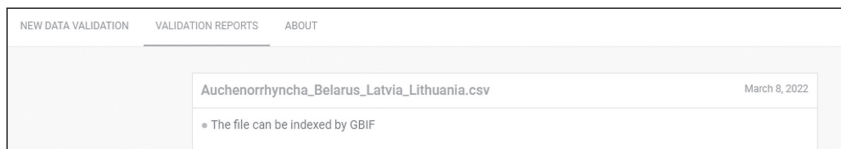


Рисунок 48

Еще одним полезным инструментом является Name Parser. Он позволяет разбить представленный для анализа таксономический список на составляющие согласно терминам Darwin Core и выполнить сравнительный анализ его соответствия с GBIF Backbone — стандартизированным таксономическим списком GBIF, включающим на момент подготовки руководства более 3,7 млн видов (рисунок 49).



Рисунок 49

Онлайн инструмент Name Parser (рисунок 50) работает с файлами формата txt, в которых приводятся построчно предлагаемые для анализа таксономические названия, файл также должен быть сохранен в кодировке UTF-8.



Рисунок 50

По результатам сопоставления генерируется таблица, на основании которой можно оценить степень совпадения анализируемого таксономического списка со списком GBIF Backbone Taxonomy с указанием ошибок в случае их наличия (рисунок 51).

scientificName	PARSER		ABOUT	
	parsed	type	genusOrAbove	specificEpithet
Anakelisia fasciata (Kirschbaum, 1868)	true	SCIENTIFIC	Anakelisia	fasciata
Anakelisia perspicillata (Boheman, 1845)	true	SCIENTIFIC	Anakelisia	perspicillata
Anoscopus albifrons (Linnaeus, 1758)	true	SCIENTIFIC	Anoscopus	albifrons
Anoscopus albiger (Germar, 1821)	true	SCIENTIFIC	Anoscopus	albiger
Anoscopus flavostriatus (Donovan, 1799)	true	SCIENTIFIC	Anoscopus	flavostriatus
Anoscopus histrionicus (Fabricius, 1794)	true	SCIENTIFIC	Anoscopus	histrionicus
Anoscopus serratulae (Fabricius, 1775)	true	SCIENTIFIC	Anoscopus	serratulae
Aphrodes bicinctus (Schrank, 1776)	true	SCIENTIFIC	Aphrodes	bicinctus
Aphrodes diminuta Ribaut, 1952	true	SCIENTIFIC	Aphrodes	diminuta
Aphrodes makarovi Zachvatkin, 1948	true	SCIENTIFIC	Aphrodes	makarovi
Aphrophora alni (Fallen, 1805)	true	SCIENTIFIC	Aphrophora	alni

Рисунок 51

Также система позволяет сгенерировать файл в формате csv для его последующего использования.

С полным списком инструментов, которые могут быть полезны при работе с GBIF и включающим в настоящее время 110 наименований, можно ознакомиться по ссылке <https://www.gbif.org/resource/search?contentType=tool>.

7. Лицензии

В соответствии с решением правления GBIF от 2014 года, при публикации наборов данных необходимо назначать для датасета лицензию Creative Commons (CC). Эта некоммерческая организация создала бесплатные для использования типовые договоры — публичные лицензии. В GBIF используется три:

CC0 — универсальная лицензия, на основании которой автор данных полностью отдает все права на пользование своими данными без каких-либо ограничений;



CC BY — CC Attribution, лицензия с указанием авторства. Лицензия позволяет пользователям данными использовать их в любых целях: распространять, перерабатывать, даже в коммерческих целях, но при условии указания авторства.



CC BY-NC — CC Attribution — Noncommercial с указанием авторства — некоммерческая. Лицензия позволяет пользователям перерабатывать, корректировать данные на некоммерческой основе с требованием сохранения авторства.

GBIF рекомендует использовать максимально свободные лицензии, способствующие открытому распространению данных.

8. Публикация набора данных через IPT (The Integrated Publishing Toolkit)

После подготовки набора данных, его очистки и подготовки к публикации, необходимо осуществить непосредственный процесс его публикации. Как уже неоднократно подчеркивалось, одним из самых распространенных вариантов является публикация через IPT, который представляет собой серверное приложение с удобным пользовательским интерфейсом.

Публиковать данные можно либо через IPT, установленный на сервере организации, либо через любой другой, список которых доступен по адресу: <https://www.gbif.org/ipt>, а также рассматривается в разделе 9 данного руководства.

Процесс установки IPT в рамках данного руководства нами не рассматривается. Более подробную информацию можно получить по адресу: <https://www.gbif.org/ipt>, где предлагается в том числе краткое руководство по установке и использованию IPT в видеоформате, также информацию можно найти по следующим ссылкам: <https://ipt.gbif.org/manual/en/ipt/2.5/>, <http://gbif.ru/iptusing>.

Независимо от того, публикуются данные через собственный IPT публикующей организации, либо через сторонние IPT, в наборе данных отображается именно публикующая организация. Существующие алгоритмы позволяют в любое время вносить корректировки в уже опубликованные наборы данных и при необходимости их можно отозвать.

Первым шагом является получение учетной записи в выбранном для работы IPT. Для этого необходимо связаться с администратором IPT, отправив ему соответствующий запрос в виде составленного в свободной форме письма. Контактные данные администратора приводятся на странице отдельного взятого IPT. Найти страницу выбранного IPT можно несколькими способами:

- ✓ через поисковую строку на сайте gbif.org;
- ✓ путем масштабирования карты расположения IPT <https://www.gbif.org/ipt> (см. раздел 9 руководства);
- ✓ через уже известный адрес расположения IPT. Например, IPT Национального узла GBIF Беларуси располагается по адресу <http://80.94.160.207:8080/ipt/>;
- ✓ связавшись с Секретариатом GBIF либо соответствующим Национальным узлом.

Также следует уточнить у администратора, зарегистрирована ли в системе данного IPT организация, которая будет выступать в качестве публикующей. Если нет, то администратору необходимо выслать коды доступа к IPT, которые были получены контактным лицом публикующей организации после ее одобрения в системе GBIF, о чем уже было сказано в разделе 3 данного руководства.

После получения учетной записи следует авторизоваться на главной странице выбранного IPT. В частности, в случае с IPT Национального узла GBIF Беларуси домашняя страница выглядит следующим образом (рисунок 52).

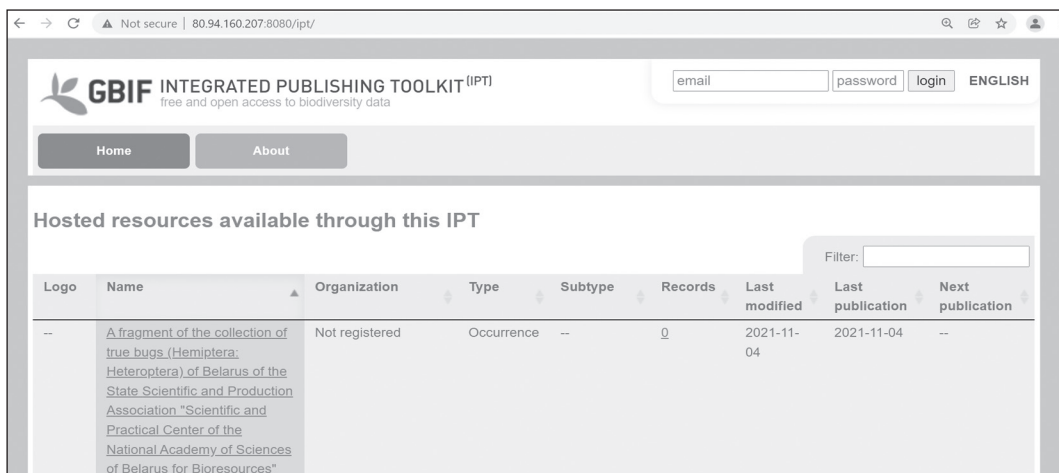


Рисунок 52

В правом верхнем углу необходимо ввести регистрационные данные (e-mail и пароль), которые вы получите от администратора IPT. В последующем можно сменить пароль в настройках аккаунта пользователя учетной записи.

Следует отметить, что даже без регистрации на IPT любому пользователю доступен список наборов данных, которые на нем находятся, даже если они еще не опубликованы в системе GBIF. При этом, подчеркнем, что редактировать можно только те наборы, которые созданы владельцем учетной записи (если нет более высокого уровня допуска).

После регистрации появляются дополнительные вкладки (рисунок 53). Вкладка «Home», по-прежнему позволяет просматривать список всех датасетов данного IPT. В появившейся после регистрации вкладке «Manage Resources» будут доступны для просмотра и редактирования, как уже было сказано, только наборы данных, созданные владельцем учетной записи. В примере на рисунке 53 создана демо-учетная запись, что позволит в конечном итоге проследить весь этап опубликования набора данных.

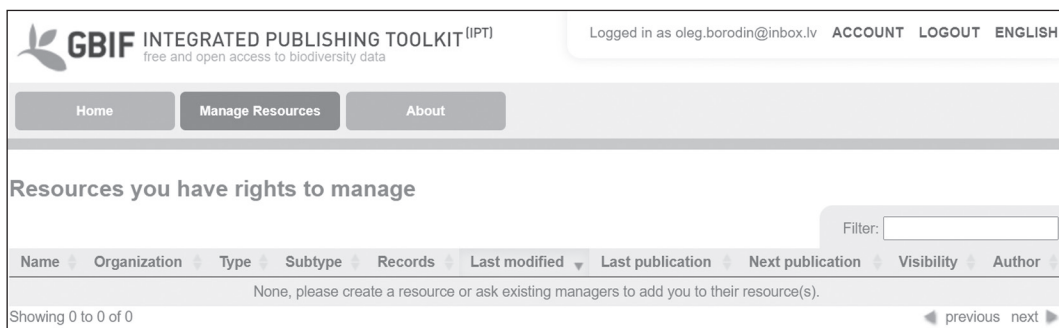


Рисунок 53

При создании нового набора данных необходимо перейти в раздел «Manage Resources», ввести краткое, черновое название создаваемого набора данных, которое будет отражаться в списке наборов данных текущего IPT (в названии должно быть использовано не менее 3 латинских символов, цифр, дефисов или подчеркиваний без пробелов).

Затем рекомендуется выбрать из выпадающего списка тип набора данных (в данном примере выбран тип набора данных «Occurrence» (Наблюдения) (см. раздел 5)) и нажать кнопку «Create» (рисунок 54).

Create New Resource

You can create a new blank resource, upload an existing resource saved as a zipped Darwin Core archive, or upload an existing IPT resource using its zipped resource configuration folder. Please refer to the User Manual for more specific instructions. A short name is required.

Shortname

Type

Import from an archived resource

Рисунок 54

При этом на сервере создается папка для вашего набора данных и его шаблон (рисунок 55). В последующем при подготовке метаданных (см. раздел 5) датасета может быть дано более полное и развернутое название. На данном этапе также возможна загрузка архива, содержащего всю необходимую информацию, в корректной форме и без ошибок, которые можно дополнительно устранить в процессе загрузки данных по рассмотренной далее схеме. Кроме того, в системе предусмотрена возможность загрузки запросов из существующих баз данных. В рамках данного руководства эти алгоритмы не рассматриваются.

Overview: [auchenorrhyncha belarus latvia lithuania](#)

This is the overview page for the *auchenorrhyncha_belarus_latvia_lithuania* resource.

Source Data Your source data files and SQL sources for generating a Darwin Core Archive.

Файл не выбран

Darwin Core Mappings Your mapping between the source data and Darwin Core terms.

Metadata Your resource metadata.

Published Versions A preview of your pending published version compared with the current version if existing.

<p>Auto-publishing</p> <p><input type="text" value="Select interval"/> <input type="button" value="ⓘ"/></p>	<table style="width: 100%; border-collapse: collapse;"> <tr> <td style="width: 50%;"></td> <td style="text-align: right; font-size: small;">Pending version</td> </tr> <tr> <td style="font-size: small;">Version</td> <td style="text-align: right;">1.0 <input type="button" value="Preview"/></td> </tr> <tr> <td style="font-size: small;">Visibility</td> <td style="text-align: right;">Private</td> </tr> <tr> <td style="font-size: small;">Data Licence</td> <td style="text-align: right;">-</td> </tr> <tr> <td style="font-size: small;">Published on</td> <td style="text-align: right;">-</td> </tr> </table>		Pending version	Version	1.0 <input type="button" value="Preview"/>	Visibility	Private	Data Licence	-	Published on	-
	Pending version										
Version	1.0 <input type="button" value="Preview"/>										
Visibility	Private										
Data Licence	-										
Published on	-										

Visibility *Private* This resource is private to managers. To make this resource available to everyone, please make it public.

Resource Managers Managers granted permission to modify this resource.

Creator Oleg Borodin, oleg.borodin@inbox.lv

Рисунок 55

На следующем этапе необходимо выбрать созданный ранее файл набора данных.

Для этого, нажав кнопку «Выберите файл», необходимо указать путь к хранящемуся файлу с расширением csv, созданный ранее, о чем было сказано в разделе 5, и нажмите кнопку «Add».

В появившемся окне (рисунок 56) проверяем предлагаемые по умолчанию параметры:

- разделитель (Field Delimiter), используемый в загружаемом файле с расширением csv, о чем было сказано в разделе 5;
- число заглавных строк (Number of Header Rows);
- формат отображения даты, например, ISO формат (год-месяц-дата (YYYY-MM-DD));
- кодировку (Character Encoding) – рекомендуется использовать кодировку UTF-8.

Added new file source >>auchenorrhyncha_belarus_latvia_lithuania<<

Resource Title auchenorrhyncha belarus latvia lithuania

Source Data

Edit your source data format

Source Name

Readable

Columns 22

File D:\ipt\resources\lithuania\sources\lithuania.txt

Size 81,3 KB

Rows 218

Modified 2022-03-07 21:33:14

Source log [Download](#)

Number of Header Rows

Field Delimiter

Field Quotes

Multi-value Delimiter

Character Encoding

Date Format

IPT Version 2.4.2-rd9bf4b [About the IPT](#) [User manual](#) [Report a bug](#) [Request new feature](#)

©2018 Global Biodiversity Information Facility.

Рисунок 56

После нажатия кнопки «Save» происходит возврат на предыдущую страницу.

На следующем этапе необходимо выполнить актуализацию полей загружаемой таблицы набора данных согласно существующим терминам стандарта Darwin Core (см. раздел 4). Для этого необходимо зайти в раздел Darwin Core Mappings, нажав соответствующую кнопку «Add» (рисунок 57).

The screenshot shows a web interface for managing Darwin Core data. It is divided into three main sections:

- Source Data:** Contains a file selection button "Выберите файл" (File not selected), a "Connect to database" button, and a "Clear" button. Below this, it displays "Last modified Mar 7, 2022" and a file entry: "auchenorrhyncha_belarus_latvia_lithuania81,3 KB, 218 rows, 22 columns. Mar 7, 2022 [file]" with an "Edit" button.
- Darwin Core Mappings:** Shows a dropdown menu for "Darwin Core Occurrence" and an "Add" button.
- Metadata:** Includes an "Edit" button.

Рисунок 57

В появившемся окне в разделе «Source data» необходимо еще раз выбрать ту версию таблицы, которая загружается в качестве ресурса набора данных. При подгрузке ранее только одного варианта таблицы в выпадающем окне она и будет представлена по умолчанию, в случае, если существует несколько версий таблиц, на данном этапе можно выбрать именно ту, которая будет являться актуальным источником информации.

После нажатия кнопки «Save» будет осуществлен переход на страницу (рисунки 58–64), где представлены термины Darwin Core, распределенные по группам согласно порядку, приводимому в разделе 4 данного руководства.

The screenshot displays the "Mapping Source Data" configuration page. At the top, a checkbox is checked with the text "Automapped 22 columns based on header names." Below this, the "Resource Title" is "auchenorrhyncha belarus latvia lithuania".

The interface is split into two columns:

- Field Index:** A list of Darwin Core terms including Record-level, MaterialSample, Occurrence, Organism, Event, Location, GeologicalContext, Identification, and Taxon. Below this is a "Field Filters" section with "Hide unmapped fields" and "Save", "Delete", "Back" buttons.
- Mapping Source Data:** A section titled "Mapping source data auchenorrhyncha_belarus_latvia_lithuania to core extension: Darwin Core Occurrence." It contains:
 - Record-level:** A table of mappings for terms like occurrenceID, Filter, dcterms:type, dcterms:modified, dcterms:language, dcterms:license, dcterms:rightsHolder, dcterms:accessRights, and dcterms:bibliographicCitation. Each row has a dropdown menu and an input field.

Рисунок 58

Если при создании таблицы все поля были заданы верно, то их сопоставление происходит автоматически, в случае наличия ошибок, появятся соответствующие предупреждения, позволяющие их визуализировать. Тогда сопоставление полей можно осуществить в ручном режиме. Для этого, просматривая все термины Darwin Core, при обнаружении выделенной цветом ошибки, верное значение можно выбрать из выпадающего списка.

В некоторых случаях допустимо введение фиксированного значения в правом поле для каждого термина. Например, если весь набор данных представлен материалами, собранными в течение одного года и т. п.

Field Index

Record-level

MaterialSample

Occurrence

Organism

Event

Location

GeologicalContext

Identification

Taxon

Field Filters

Hide unmapped fields

MaterialSample

Occurrence

Source Sample: | | | |

Translation:

Source Sample: Borodin O.I. | Borodin O.I. | Borodin O.I. | Borodin O.I. | Borodin O.I.

Translation:

Source Sample: 1f | 1m | 1 adult without abdomen | 3m | 3m, 1f

Translation:

Рисунок 59

Field Index

Record-level

MaterialSample

Occurrence

Organism

Event

Location

GeologicalContext

Identification

Taxon

Field Filters

Hide unmapped fields

Event

Source Sample: 2005 | 2004 | 2002 | 2013 | 2013

Translation:

Source Sample: 8 | 8 | 9 | 9 | 9

Translation:

Source Sample: 15 | 5 | 6 | 28 | 29

Translation:

Рисунок 60

40

	<input type="checkbox"/> minimumElevationInMeters	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> maximumElevationInMeters	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimDepth	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> minimumDepthInMeters	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> maximumDepthInMeters	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> minimumDistanceAboveSurface	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> maximumDistanceAboveSurface	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> locationAccordingTo	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> locationRemarks	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimCoordinates	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimLatitude	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimLongitude	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimCoordinateSystem	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> verbatimSRS	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> decimalLatitude	<input type="text" value="decimalLatitude"/>	<input type="text"/>
	Source Sample: 51.936376 51.884368 54.267174 53.946148 53.938178 Translation: <input type="button" value="Add"/>		
	<input type="checkbox"/> decimalLongitude	<input type="text" value="decimalLongitude"/>	<input type="text"/>
	Source Sample: 26.9044104 28.459298 27.035176 27.127104 27.385924 Translation: <input type="button" value="Add"/>		
	<input type="checkbox"/> geodeticDatum	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> coordinateUncertaintyInMeters	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> coordinatePrecision	<input type="text"/>	<input type="text"/>

Field Index

Record-level

MaterialSample

Occurrence

Organism

Event

Location

GeologicalContext

Identification

Taxon

Field Filters

Hide unmapped fields

Рисунок 61

	<input type="checkbox"/> taxonID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> scientificNameID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> acceptedNameUsageID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> parentNameUsageID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> originalNameUsageID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> nameAccordingToID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> namePublishedInID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> taxonConceptID	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> scientificName	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> acceptedNameUsage	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> parentNameUsage	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> originalNameUsage	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> nameAccordingTo	<input type="text"/>	<input type="text"/>
	<input type="checkbox"/> namePublishedIn	<input type="text"/>	<input type="text"/>

Field Index

Record-level

MaterialSample

Occurrence

Organism

Event

Location

GeologicalContext

Identification

Рисунок 62

Field Index Record-level MaterialSample Occurrence Organism Event Location GeologicalContext Identification Taxon Field Filters Hide unmapped fields <input type="button" value="Save"/> <input type="button" value="Delete"/> <input type="button" value="Back"/>	<input type="text" value="namePublishedInYear"/>
	<input type="text" value="higherClassification"/>
	<input type="text" value="kingdom"/> kingdom Source Sample: <i>Animalia</i> <i>Animalia</i> <i>Animalia</i> <i>Animalia</i> <i>Animalia</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="phylum"/> phylum Source Sample: <i>Arthropoda</i> <i>Arthropoda</i> <i>Arthropoda</i> <i>Arthropoda</i> <i>Arthropoda</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="class"/> class Source Sample: <i>Ectognatha</i> <i>Ectognatha</i> <i>Ectognatha</i> <i>Ectognatha</i> <i>Ectognatha</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="order"/> order Source Sample: <i>Hemiptera</i> <i>Hemiptera</i> <i>Hemiptera</i> <i>Hemiptera</i> <i>Hemiptera</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="family"/> family Source Sample: <i>Aphalaridae</i> <i>Aphalaridae</i> <i>Aphalaridae</i> <i>Aphalaridae</i> <i>Aphalaridae</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="genus"/> genus Source Sample: <i>Aphalara</i> <i>Aphalara</i> <i>Aphalara</i> <i>Aphalara</i> <i>Aphalara</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="subgenus"/>
	<input type="text" value="specificEpithet"/> specificEpithet Source Sample: <i>avicularis</i> <i>freji</i> <i>affinis</i> <i>freji</i> <i>freji</i> Translation: <input type="button" value="Add"/>
<input type="text" value="infraspecificEpithet"/>	
<input type="text" value="taxonRank"/>	

Рисунок 63

MaterialSample Occurrence Organism Event Location GeologicalContext Identification Taxon Field Filters Hide unmapped fields <input type="button" value="Save"/> <input type="button" value="Delete"/> <input type="button" value="Back"/>	<input type="text" value="verbatimTaxonRank"/>
	<input type="text" value="scientificNameAuthorship"/> scientificNameAuthorship Source Sample: <i>Ossiannilsson, 1981</i> <i>Burckhardt & Lauterer, 1997</i> <i>(Zetterstedt, 1828)</i> <i>Burckhardt & Lauterer, 1997</i> <i>Burckhardt & Lauterer, 1997</i> Translation: <input type="button" value="Add"/>
	<input type="text" value="vernacularName"/>
	<input type="text" value="nomenclaturalCode"/>
	<input type="text" value="taxonomicStatus"/>
	<input type="text" value="nomenclaturalStatus"/>
	<input type="text" value="taxonRemarks"/>
	<input type="button" value="Save"/> <input type="button" value="Delete"/> <input type="button" value="Back"/>

[IPT Version 2.4.2-rd9bfb4b](#)
[About the IPT](#)
[User manual](#)
[Report a bug](#)
[Request new feature](#)

©2018 Global Biodiversity Information Facility.

Рисунок 64

В случае наличия проблемных полей, их список будет отображаться в конце списка терминов Darwin Core (например, рисунок 65), что упрощает их поиск и корректировку.

taxonomicStatus	①	<input type="text"/>	<input type="text"/>
nomenclaturalStatus	①	<input type="text"/>	<input type="text"/>
taxonRemarks	①	<input type="text"/>	<input type="text"/>

Unmapped columns

The following columns of the source are not assigned to the properties of this mapping:

- Field no.
- At
- Reserve
- Substrate
- Collector name
- Reference
- Title(Link to material)

Рисунок 65

После проверки соответствия всех полей терминам Darwin Core необходимо нажать кнопку «Save» (рисунок 64). Произойдет переход в верхнюю часть списка, где необходимо нажать на ссылку названия набора данных (рисунок 66).

Home Manage Resources Administration About	
Resource Title	<u>auchenorrhyncha belarus latvia lithuania</u>
Field Index	① Mapping Source Data
Record-level	Mapping source data <u>auchenorrhyncha_belarus_latvia_lithuania</u> to core extension: Darwin Core Occurrence .
MaterialSample	① occurrenceID <input type="text" value="occurrenceID"/>
Occurrence	Source Sample: 1 2 3 4 5
Organism	① Filter <input type="text" value="AfterTranslation"/>
Event	
Location	
GeologicalContext	
Identification	① dcterms:type <input type="text"/>
Taxon	① dcterms:modified <input type="text"/>
Field Filters	① dcterms:language <input type="text"/>
Hide unmapped fields	① dcterms:license <input type="text"/>
<input type="button" value="Save"/> <input type="button" value="Delete"/> <input type="button" value="Back"/>	① dcterms:rightsHolder <input type="text"/>

Рисунок 66

После возвращения на страницу набора данных необходимо провести описание набора данных, заполнив форму в разделе «Metadata». Для этого следует нажать кнопку «Edit» в соответствующем разделе (рисунок 67).

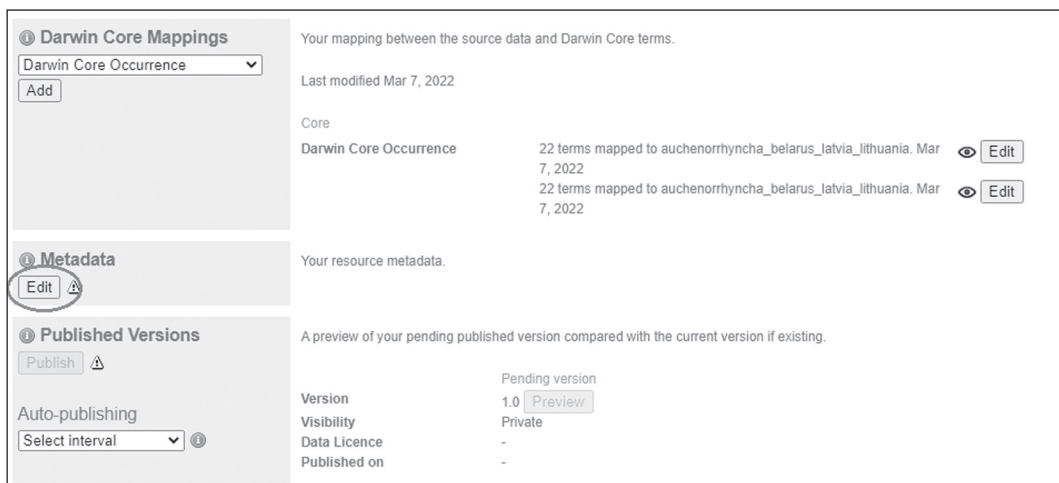


Рисунок 67

В открывшемся окне необходимо указать всю запрашиваемую информацию (рисунок 68). В обязательном порядке необходимо заполнять поля, помеченные звездочкой. Однако, следует понимать, что на данном этапе чем более полноценно будет заполнена форма, тем более детально вы опишите свой набор данных и сделаете его максимально доступным для пользователей. Также рекомендуется заполнять форму на английском языке.

В правой части окна располагается дерево формы, позволяющее оперативно перемещаться между ее разделами. Последовательно их рассмотрим.

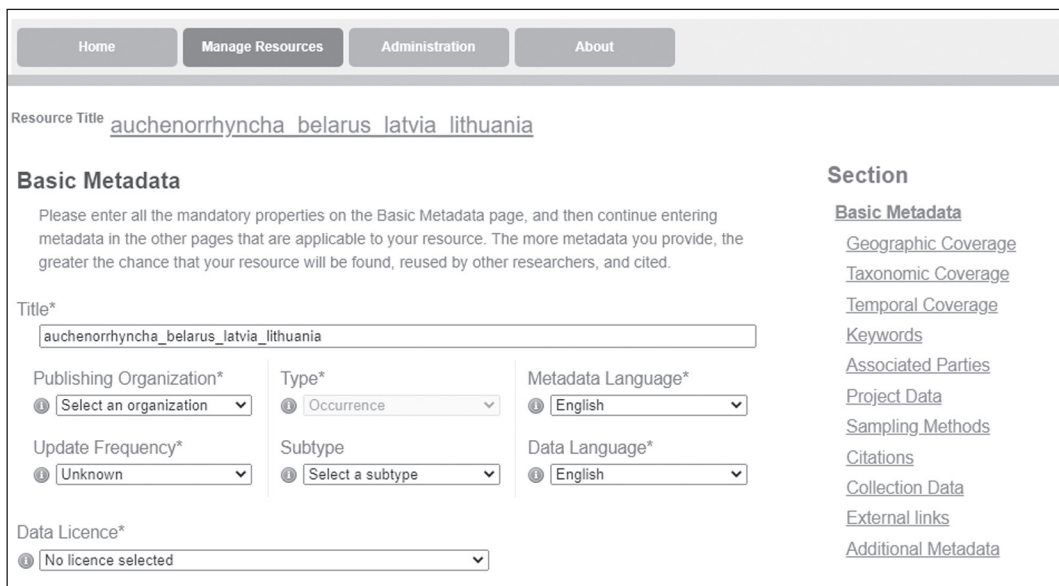


Рисунок 68

Первое открывшееся окно содержит базовую информацию о наборе данных. В строке «Title» по умолчанию приводится краткое название набора данных, которое было предложено вами в самом начале работы (рисунок 68). На данном этапе здесь можно привести его полное название (рисунок 69).

Home Manage Resources Administration About

Resource Title auchenorrhyncha belarus latvia lithuania

Basic Metadata

Please enter all the mandatory properties on the Basic Metadata page, and then continue entering metadata in the other pages that are applicable to your resource. The more metadata you provide, the greater the chance that your resource will be found, reused by other researchers, and cited.

Title*

Publishing Organization*

Type*

Metadata Language*

Update Frequency*

Subtype

Data Language*

Data Licence*

This work is licensed under a [Creative Commons Attribution \(CC-BY\) 4.0 License](#).

You should only apply a licence to your own work unless you have the necessary rights to apply it to another person's.

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 69

Также на данном этапе необходимо выбрать название публикующей организации, от имени которой публикуется набор данных (см. раздел 3). Также необходимо подтвердить либо откорректировать значения в строках «Metadata Language» и «Data Language». По умолчанию предлагается английский язык в качестве языка описания набора данных и данных в таблице. Конечно, в конечном итоге выбор за владельцем данных, но согласно существующей практике набор данных и его описание, опубликованные на английском языке, более доступны широкому кругу пользователей.

В строке «Type» по умолчанию установлен ранее выбранный тип наборов данных «Occurrence» — наблюдение. В строке «Update Frequency» предлагается выбрать вариант, с какой периодичностью будет проводиться обновление текущего набора данных. По умолчанию предлагается вариант «Unknown». В строке «subtype» в выбранном типе наборов данных в выпадающем списке предлагается вариант «specimen» (экземпляр) или «observation» (наблюдение). Учитывая, что в загруженном наборе данных приводятся сведения для находящихся в коллекции экземпляров, выбирается вариант «specimen». Вариант «observation» может быть выбран, в частности, в случае формирования набора данных на основании полевого дневника наблюдений, например, за птицами.

Наконец, на данном этапе необходимо выбрать вариант лицензии, согласно которой вы разрешаете использовать созданный набор данных. О типах лицензий более подробно можно посмотреть в разделе 7.

На этой же странице необходимо привести как можно более подробное описание публикуемого набора данных в разделе «Description». В приводимом тестовом примере дана краткая информация (рисунок 70).

The screenshot displays the IPT (Integrated Publishing Tool) interface. At the top, the 'Description*' section is active, featuring a text area with the following text: "Dataset contains information on the distribution of Cicadomorpha and Fulgoromorpha, living in the territories of Belarus, Latvia and Lithuania. In addition to the original collection data, the dataset contains published information." A link "[Remove this Paragraph]" is located to the right of the text area. Below the text area is a link "[Add new Paragraph]".

Below the description section are three other sections, each with a link to add a new entry:

- Resource Contacts*** with a link "Add new resource contact".
- Resource Creators*** with a link "Add new resource creator".
- Metadata Providers*** with a link "Add new metadata provider".

At the bottom of the form area, there are "Save" and "Cancel" buttons. The footer of the page contains the text "IPT Version 2.4.2-rd9bfb4b" and several links: "About the IPT", "User manual", "Report a bug", and "Request new feature". The copyright notice "©2018 Global Biodiversity Information Facility." is located at the very bottom.

Рисунок 70

Также на этой же странице необходимо предоставить контактную информацию (рисунки 70, 71). В разделе «Resource Contacts» приводится информация о специалистах либо организациях, которые курируют набор данных и где можно получить более подробную информацию о нем. В разделе «Resource Creators» информация о специалистах, которые создали публикуемый набор данных, наконец в разделе «Metadata Providers» приводятся сведения о специалистах, подготовивших описание набора данных. Учитывая, что в данной работе могут принимать участие различные специалисты, сведения о каждом из них последовательно добавляются путем нажатия ссылки «Add new ...» в соответствующем разделе.

Resource Contacts*

[[Copy details from resource contact](#)] [[Remove this resource contact](#)]

First Name	Last Name*
<input type="text"/>	<input type="text"/>
Position*	Organization*
<input type="text"/>	<input type="text"/>
Address	City
<input type="text"/>	<input type="text"/>
State/Province	Country
<input type="text"/>	<input type="text" value="Select a country, territory, or island"/>
Postal Code	Phone
<input type="text"/>	<input type="text"/>
Email	Home Page
<input type="text"/>	<input type="text"/>
Personnel Directory	Personnel Identifier
<input type="text" value="Select a directory"/>	<input type="text"/>

[Add new resource contact](#)

Рисунок 71

После нажатия кнопки «Save» в нижней части страницы будет произведен переход на следующую страницу «Geographic Coverage» (Географический охват) (рисунок 72).

В случае наборов данных, охватывающих всю планету (например, наборы данных крупных музеев либо глобальных проектов) можно оставить предлагаемую по умолчанию опцию «Set global coverage», если же набор охватывает определенные географические регионы можно либо выбрать его на карте, либо, убрав маркер, указать в появившихся полях (рисунок 73) крайние точки.

Необходимо также привести описание географического охвата, указав, например, страны, физико-географические районы, территории Национальных парков и т. д.

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Geographic Coverage

Drag the markers or fill in the fields to set the geographic bounding box of the area covered by the resource.



Set global coverage

Description*

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)**
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 72

Set global coverage

West* East*

South* North*

Description*

Entire territory of Belarus, Latvia and Lithuania

IPT Version 2.4.2-rd9bfb4b [About the IPT](#) [User manual](#) [Report a bug](#) [Request new feature](#)

©2018 Global Biodiversity Information Facility.

Рисунок 73

После нажатия кнопки «Save» будет произведен переход на следующий раздел «Taxonomic Coverage» (Таксономический охват) (рисунок 74). Раздел не является обязательным для заполнения. Тем не менее, предоставление данной информации (рисунок 75) сделает его более доступным для целевой аудитории.

Geographic Coverage section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Taxonomic Coverage

Please enter metadata about the taxonomic areas covered by the resource.

[Add new taxonomic coverage](#)

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)**
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 74

Geographic Coverage section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Taxonomic Coverage

Please enter metadata about the taxonomic areas covered by the resource.

[\[Remove this taxonomic coverage \]](#)

Description

[Add several taxa](#)

[Add new taxon](#)

[Add new taxonomic coverage](#)

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)**
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 75

В случае наборов данных, охватывающих широкий спектр таксономических групп, на данном этапе можно привести более детальное их описание.

Следующим разделом будет «Temporal Coverage» (временной охват). Здесь приводятся сведения о временном диапазоне. Раздел также не является обязательным для заполнения (рисунок 76).

Taxonomic Coverage section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)**
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Temporal Coverage

Please enter metadata about the time periods covered by the resource. First select the type of time period, then fill in the form fields that appear.

Temporal Coverage Type [\[Remove this temporal coverage \]](#)

Single Date

Single Date

[Add new temporal coverage](#)

Рисунок 76

При заполнении раздела можно выбрать несколько вариантов отображения информации (рисунк 77).

Temporal Coverage

Please enter metadata about the time periods covered by the resource. First select the type of time period, then fill in the form fields that appear.

Temporal Coverage Type [\[Remove this temporal coverage \]](#)

Single Date

Single Date

Formation Period

Date Range

Living Time Period

[Add new temporal coverage](#)

Рисунок 77

Single Date – подразумевает, что будет указан единственный день, в течение которого была собрана информация для данного датасета.

Formation Period – данный тип отображения временного охвата предназначен для отображения его текстовой форме, например, Векторианская эпоха, с 1970 года, 2010 – настоящее время и т. д.

Data Range – дает возможность четко указать временной период (рисунок 78).

Temporal Coverage Type [\[Remove this temporal coverage \]](#)

Date Range ▼

Start Date End Date

Рисунок 78

Наконец, закладка Living Time Period используется для описания временного периода, в течение которого коллекционный материал был жив, в частности, в случае описания палеонтологических коллекций, в таком случае указывается, например, плейстоцен и т. п.

Следующим разделом в описании набора данных будут ключевые слова (рисунок 79).

Temporal Coverage section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Keywords

Please enter sets of keywords for the resource, each with a thesaurus or controlled vocabulary. [\[Remove this keywords \]](#)

Thesaurus/Vocabulary*

Keyword List*

[\[Remove this keywords \]](#)

Thesaurus/Vocabulary*

Keyword List*

[Add new keywords](#)

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- Keywords**
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 79

На данном этапе можно составить список ключевых слов, которые будут характеризовать ваш набор данных. Можно воспользоваться списком слов, представленных в словарях, ссылки на которые даются по умолчанию в соответствующих строках. Если слово не выбрано из какого-либо словаря, необходимо в данной строке указать значение п/а.

В следующем разделе «Associated Parties» (Ассоциированные Стороны) (рисунок 80) можно привести дополнительную информацию о всех, кто имеет отношение к данному набору данных, с указанием соответствующих ролей каждого из участников (рисунок 81).

Keywords section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Associated Parties

Parties associated with this resource (e.g., the hosting institution).

[[Copy details from resource contact](#)] [[Remove this associated party](#)]

First Name	Last Name*
<input type="text"/>	<input type="text"/>
Position*	Organization*
<input type="text"/>	<input type="text"/>
Address	City
<input type="text"/>	<input type="text"/>
State/Province	Country
<input type="text"/>	<input type="text" value="Select a country, territory, or island"/>
Postal Code	Phone
<input type="text"/>	<input type="text"/>
Email	Home Page
<input type="text"/>	<input type="text"/>
Personnel Directory	Personnel Identifier
<input type="text" value="Select a directory"/>	<input type="text"/>
Role	
<input type="text" value="Select a role"/>	

[Add new associated party](#)

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)**
- [Project Data](#)
- [Sampling Methods](#)
- [Citations](#)
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 80

На следующем этапе (рисунок 82) можно привести детальное описание проекта, в рамках которого был получен публикуемый набор данных. Здесь может быть указан номер проекта, приведено его описание, спонсоры, временной и географический охват и любая другая информация, которая по мнению авторов набора данных может быть отображена.

То же самое касается и данных о ассоциированных с проектом исполнителей, имеющих отношение к набору данных. Для данной информации предусмотрен соответствующий блок.

На следующем этапе «Sampling Methods» (Методы Сбора) можно перечислить и описать все методы, которые были использованы для получения информации, предоставленной в публикуемом наборе данных (рисунок 83).

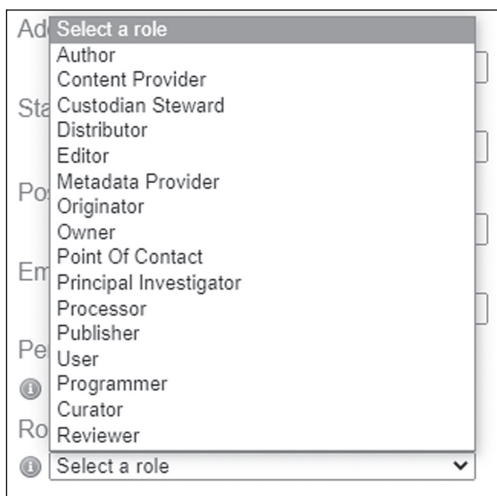


Рисунок 81

Project Data

Please enter metadata about the project under which the data in this resource were produced.

Title*

Identifier

Description

Funding

Study Area Description

Design Description

Project Personnel* _____

[Add new personnel](#)

[Geographic Coverage](#)

[Taxonomic Coverage](#)

[Temporal Coverage](#)

[Keywords](#)

[Associated Parties](#)

Project Data

[Sampling Methods](#)

[Citations](#)

[Collection Data](#)

[External links](#)

[Additional Metadata](#)

Рисунок 82

Resource Title	Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania	Section
Sampling Methods		Basic Metadata
Please enter metadata about the sampling methods used for the data represented by the resource.		Geographic Coverage
Study Extent*	<input type="text"/>	Taxonomic Coverage
Sampling Description*	<input type="text"/>	Temporal Coverage
Quality Control	<input type="text"/>	Keywords
Step Description*	<input type="text"/>	Associated Parties
Add new method step		Project Data
<input type="button" value="Save"/> <input type="button" value="Cancel"/>		Sampling Methods
		Citations
		Collection Data
		External links
		Additional Metadata

Рисунок 83

Информацию рекомендовано предоставить в следующих блоках:

Study Extent – масштаб исследований, поле, позволяющее отобразить репрезентативность выборки, полноту географической, временной выборки и т. д.

Sampling Description – описание выборки, поле, позволяющее отобразить в текстовой форме описание процедуры отбора проб, по сути этот блок соответствует традиционным разделам с описанием методик исследований в научных статьях.

Quality Control – контроль качества, поле, в котором описывается алгоритм, обеспечивающий управление и оценку качества данных.

Step Description – пошаговое описание, поле, содержащее любую информацию об используемых методах, применяемом в программном обеспечении и т. д.

В следующем разделе генерируется цитата, которая отображается при описании набора данных и на которую ссылаются пользователи (рисунок 84).

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

Citations

Please enter how your resource should be cited. To ensure your resource is reliably and consistently cited, let the IPT auto-generate the citation for you. An optional bibliography can also be entered, i.e., citations of other resources related to or used in the creation of this resource.

BE AWARE: Free-text citations get overwritten on the GBIF.org dataset page – learn more here and here.

[Auto-generation is On – Turn off]

Resource Citation*

```

Borodin O (2022): Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania. v1. The Scientific and Practical Center of the National academy of Sciences of Belarus for Biological Resources. Dataset/Occurrence.
http://80.94.160.207:8080/ipt/resource?r=auchenorrhyncha_belarus_latvia_lithuania&v=1.0

```

Resource Citation Identifier

Bibliographic Citations

[Add new bibliographic citation](#)

Section

- [Basic Metadata](#)
- [Geographic Coverage](#)
- [Taxonomic Coverage](#)
- [Temporal Coverage](#)
- [Keywords](#)
- [Associated Parties](#)
- [Project Data](#)
- [Sampling Methods](#)
- Citations**
- [Collection Data](#)
- [External links](#)
- [Additional Metadata](#)

Рисунок 84

Цитата может быть сформулирована и автором набора данных. На данной странице приведены ссылки на более подробные инструкции по ее составлению.

На следующих этапах можно привести более подробную информацию о коллекциях, где хранится материал (рисунок 85).

Collection Data

Please enter the collection metadata for the resource.

Collections

[[Remove this collection](#)]

Collection Name*

Collection Identifier

Parent Collection Identifier

[Add new collection](#)

Рисунок 85

Методах, используемых при хранении коллекционного материала (рисунок 86).

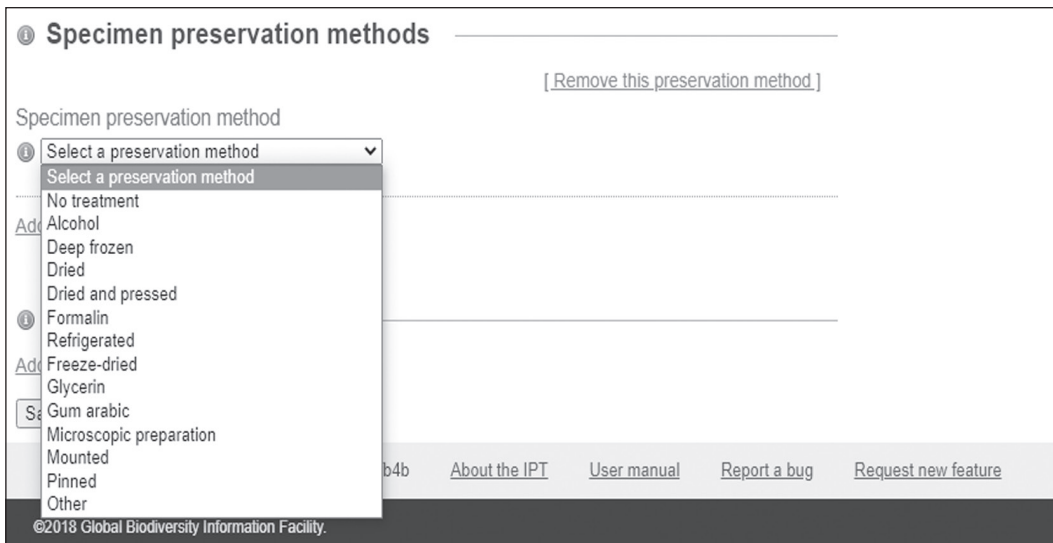


Рисунок 86

Как и во всех предыдущих, аналогичных примерах, в случае использования нескольких методов хранения, каждый из них может быть описан отдельно.

Также в этом же разделе еще можно указать сведения о количестве элементов коллекции, которые охвачены публикуемым набором данных (например, количество коллекционных шкафов, гербарных папок и т. д.) (рисунок 87).

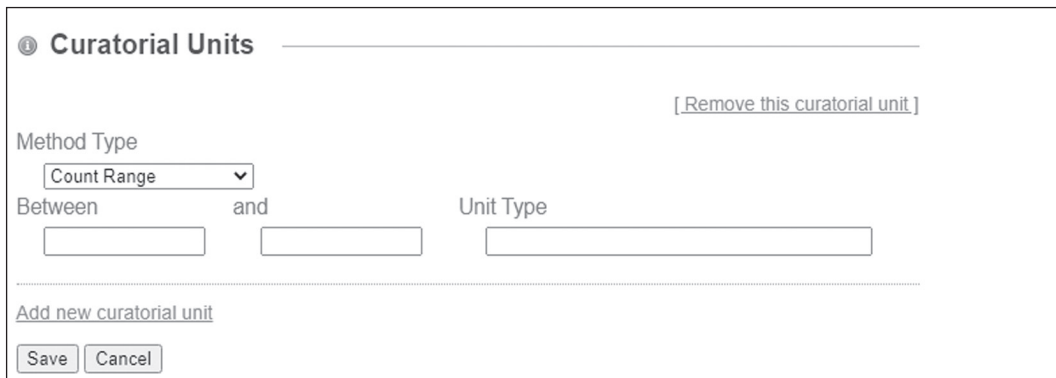


Рисунок 87

Следующий раздел «External Links» (Внешние ссылки) позволяет представить дополнительную информацию в виде внешнего ресурса, описывающего публикуемый набор данных или предоставляющий дополнительную информацию относительно него (рисунок 88).

Collection Data section successfully saved!

Resource Title Fulgoromorpha & Cicadomorpha (Hemiptera) of Belarus, Latvia and Lithuania

External Links
 External links to your resource homepage or other available formats (e.g., database dumps, spreadsheets, nexus, linked data, etc.).
 Resource Homepage

Other Data Formats _____
[Add new external link](#)

Section
[Basic Metadata](#)
[Geographic Coverage](#)
[Taxonomic Coverage](#)
[Temporal Coverage](#)
[Keywords](#)
[Associated Parties](#)
[Project Data](#)
[Sampling Methods](#)
[Citations](#)
[Collection Data](#)
External links
[Additional Metadata](#)

IPT Version 2.4.2-rd9fb4b [About the IPT](#) [User manual](#) [Report a bug](#) [Request new feature](#)

©2018 Global Biodiversity Information Facility.

Рисунок 88

И последняя страница раздела Metadata, «Additional Metadata» (Дополнительные метаданные), позволяет добавить любую дополнительную информацию (рисунок 89).

Additional Metadata
 Please enter the additional metadata for the resource.

Date Created Resource logo URL

Date Last Published Файл ...выбран

Purpose

Maintenance Description

Additional Information

Alternative Identifiers _____

Рисунок 89

После нажатия кнопки «Save» в данном разделе происходит возврат на первую страницу «Basic Metadata» (рисунок 69), что позволяет еще раз просмотреть всю внесенную информацию.

Для продолжения работы необходимо вновь нажать на ссылку с названием набора данных, что позволит вернуться в главное меню. На данном этапе необходимо нажать кнопку «Public» в разделе «Visibility» (рисунок 90).

The screenshot displays a web interface for managing Darwin Core data. It is organized into several sections:

- Source Data:** Includes a file selection area with a 'Connect to database' button and a 'Clear' button. Below, it lists a data file: 'auchenorrhyncha_belarus_latvia_lithuania2, 1 MB, 5,672 rows, 22 columns. Mar 8, 2022 [file]' with an 'Edit' button.
- Darwin Core Mappings:** Shows a dropdown menu set to 'Darwin Core Occurrence' and an 'Add' button. It lists mappings for 'Darwin Core Occurrence' with '22 terms mapped to auchenorrhyncha_belarus_latvia_lithuania. Mar 7, 2022', each with an 'Edit' button.
- Metadata:** Features an 'Edit' button.
- Published Versions:** Contains a 'Publish' button and an 'Auto-publishing' dropdown set to 'Select interval'. A comparison table shows the 'Pending version' (1.0) with 'Preview', 'Private' visibility, and 'CC-BY 4.0' license.
- Visibility Private:** A 'Public' button is circled in red. The text below states: 'This resource is private to managers. To make this resource available to everyone, please make it public.'
- Resource Managers:** Includes an 'Add' button and a list of managers with a 'Delete' button at the bottom.

Рисунок 90

Статус сменится с Частного на Общедоступный, Публичный (рисунок 91).

Также на этой странице можно добавить список менеджеров набора данных, которые могут редактировать этот ресурс, и обозначить планируемую периодичность обновления набора данных в случае автоматизации этого процесса.

Любые изменения, касающиеся внесения правок, загрузки новых версий таблиц с набором данных, все отображается на данной странице в качестве истории набора данных.

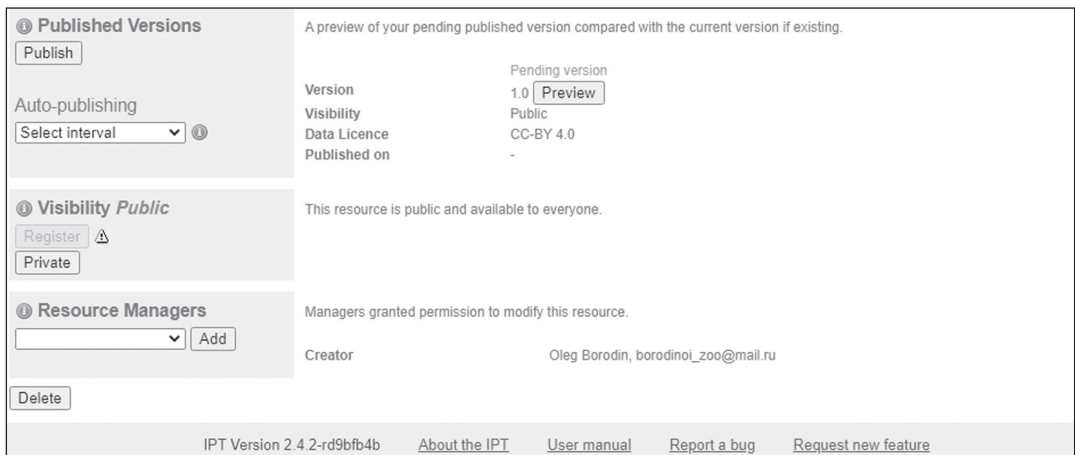


Рисунок 91

На следующем этапе необходимо нажать кнопку «Published Versions». При это формируется архив Darwin Core, который содержит все занесенные и загруженные вами данные и указанные мета-данные. В связи с тем, что вы в результате проведенных действий сформировали новый ресурс, который пока еще не зарегистрирован в сети GBIF, то этот архив пока доступен только вам и администратору IPT.

После формирования архива, если система не показала наличие ошибок, необходимо провести его регистрацию в системе GBIF. Для этого необходимо нажать кнопку «Register». Она находится под кнопкой «Publish» (рисунок 92).

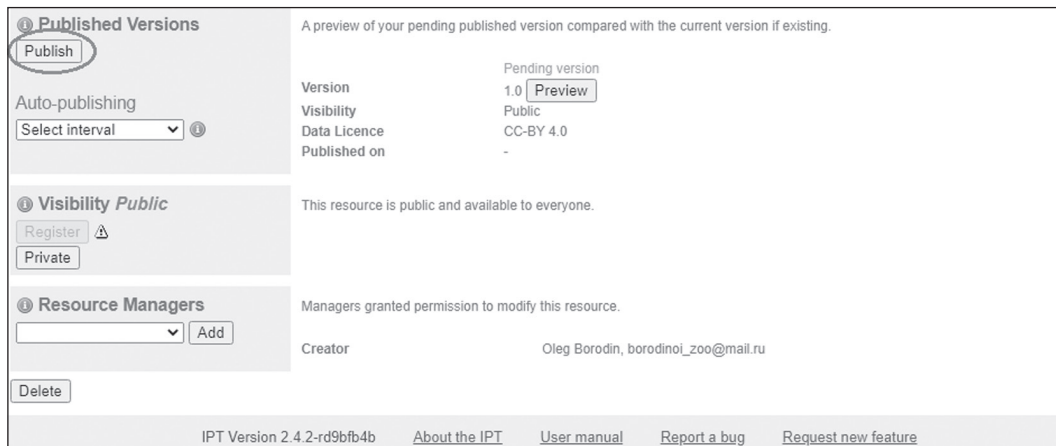


Рисунок 92

На данном этапе происходит регистрация набора данных в сети GBIF, а также индексация записей, которые в нем содержатся. В ряде случаев этот процесс может занять несколько минут.

9. Существующие IPT

На момент подготовки руководства в системе GBIF было зарегистрировано 303 IPT, располагающихся на территории 74 стран (рисунок 93).

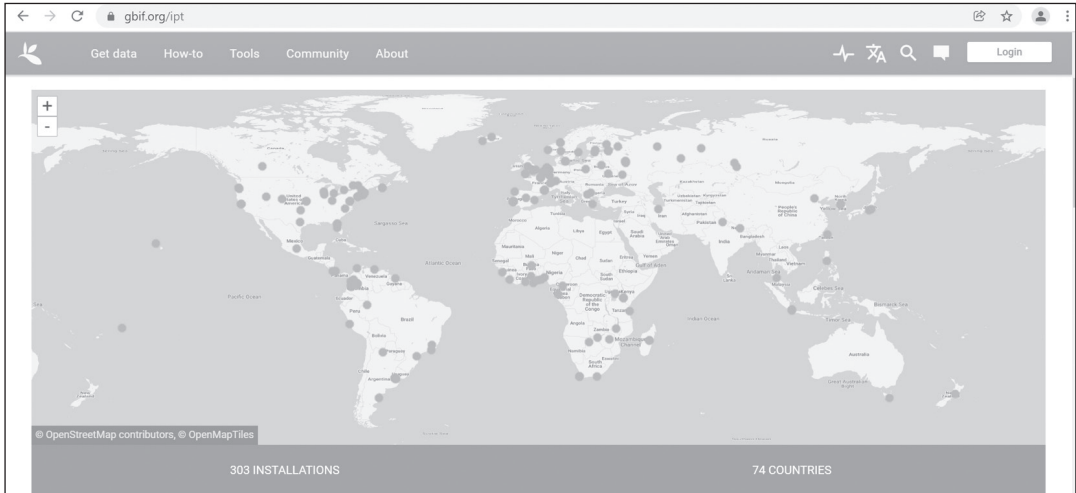


Рисунок 93

Посмотреть список можно в разделе «Tools» в соответствующей вкладке (рисунок 94).

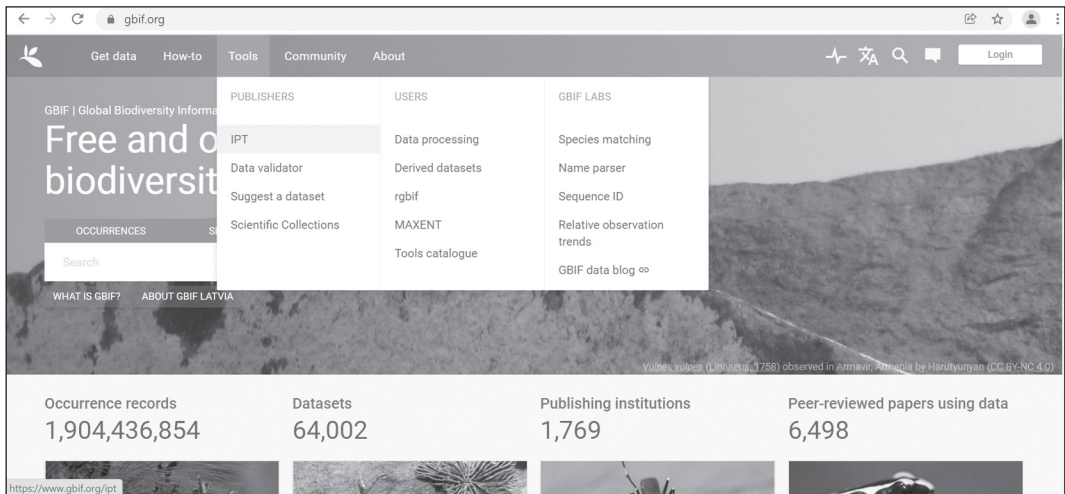


Рисунок 94

В Беларуси в настоящее время функционирует пока только один сервер IPT, ассоциированный с национальным узлом GBIF при ГНПО «НПЦ НАН Беларуси по биоресурсам» и располагающийся по адресу: <http://80.94.160.207:8080/ipt/>

Список рекомендуемых источников

1. Ваганов, А.В. Современное состояние и перспективы развития цифровой коллекции «Виртуальный гербарий АЛТВ» / А.В. Ваганов, К.В. Медведева // Проблемы ботаники Южной Сибири и Монголии. — 2020. — № 19-1. — С. 260—263. DOI: 10.14258/pbssm.2020051
2. Иванова, Н.В. Использование стандарта Darwin Core для обобщения данных о распространении охраняемых видов (на примере лишайника *Lobaria pulmonaria*) // Научные основы устойчивого управления лесами : материалы II Всерос. науч. конф. (с международным участием) / Н.В. Иванова, М.П. Шашков, В.Н. Шанин. — М. : ЦЭПЛ РАН, 2016. — С. 98—99.
3. Серегин, А.П. Гербарий московского университета (MW) сегодня : фонды, онлайн доступ и научная работа / А.П. Серегин // Ботанический журнал. — 2017. — Т. 102, № 3. — С. 281—308. DOI: 10.1134/S0006813617030012
4. Филиппова, Н.В. Информатика биоразнообразия : мировые тенденции, состояние дел в России и развитие направления в Ханты-Мансийском автономном округе // Динамика окружающей среды и глобальные изменения климата / Н.В. Филиппова, И.В. Филиппов, Д.С. Щигель [и др.]. — 2017. Т. 8, №. 2. — С. 46—56.
5. Chapman, A.D. 2005. Principles of Data Quality, version 1.0. Report for the Global Biodiversity Information Facility / A.D. Chapman. Copenhagen. — 60 p.
6. Chapman, A.D. 2005. Principles and Methods of Data Cleaning — Primary Species and Species Occurrence Data, version 1.0. Report for the Global Biodiversity Information Facility / A.D. Chapman. Copenhagen. — 75 p.
7. Endresen, D. (2019) BioDATA : Biodiversity data mobilisation and data publication training in Eurasia / D. Endresen et al. Biodiversity Information Science and Standards 3: e37543. <https://doi.org/10.3897/biss.3.37543>
8. Filippova, N. (2019) Sampling event dataset on five-year observations of macrofungi fruit bodies in raised bogs, Western Siberia, Russia / N. Filippova, E. Lapshina. Biodiversity Data Journal 7: e35674. DOI: 10.3897/BDJ.7.e35674
9. Ivanova N.V., Shashkov M.P. Application of Darwin Core standard to summarize the data of occurrences of the protected species (the example of the lichen *Lobaria pulmonaria*) // Математическая биология и биоинформатика : VI Междунар. конф., г. Пущино, 16—21 октября 2016 г. : доклады / N.V. Ivanova, M.P. Shashkov. — М. : МАКС Пресс, 2016. — С. 169—170.
10. Robertson, T. The GBIF Integrated Publishing Toolkit : Facilitating the Efficient Publishing of Biodiversity Data on the Internet // Plos One, 2014 / T. Robertson et al. <https://doi.org/10.1371/journal.pone.0102623>

Содержание

Введение	3
1. Общий алгоритм подготовки наборов данных для их публикации через GBIF	4
2. Регистрация на сайте в качестве пользователя gbif.org	5
3. Регистрация организации в качестве публикующей	7
4. Стандартизация наборов данных	14
5. Типы наборов данных, публикуемых в GBIF, и их структура	24
6. Качество данных	30
7. Лицензии	34
8. Публикация набора данных через IPT (The Integrated Publishing Toolkit)	35
9. Существующие IPT	60
Список рекомендуемых источников	61

Для записей

Производственно-практическое издание

**КРАТКОЕ РУКОВОДСТВО ПО ПОДГОТОВКЕ НАБОРОВ ДАННЫХ
ДЛЯ ПУБЛИКАЦИИ ЧЕРЕЗ ГЛОБАЛЬНЫЙ ИНФОРМАЦИОННЫЙ
ФОНД ПО БИОРАЗНООБРАЗИЮ (GBIF)**

Составитель **Бородин** Олег Игоревич

Компьютерная верстка *Н.А. Красковской*

Корректор *А.А. Чуйкова*

Подписано в печать 17.03.2022. Формат 70×90^{1/16}. Бумага офсетная. Печать офсетная.
Усл. печ. л. 4,66. Уч.-изд. л. 2,03. Тираж 100 экз. Заказ

Издательское частное унитарное предприятие «Экоперспектива».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 1/199 от 20.02.2014.
Ул. Станиславского, 23, оф. 13, 220033, г. Минск.

Общество с ограниченной ответственностью «НАВИТЕХ».
Свидетельство о государственной регистрации издателя, изготовителя,
распространителя печатных изданий № 2/194 от 23.02.2017.
Ул. Бабушкина, 6а, 220024, г. Минск.