Ускорение исследований биоразнообразия с помощью ДНК-баркодинга, коллекционных данных и данных наблюдений

Секретариат GBIF

Версия 2, April 2023

Содержание

Описание курса

Цель

При прохождении курса вы научитесь использовать штрихкоды ДНК, данные коллекций и наблюдений для решения исследовательских вопросов в области биоразнообразия. В программе комбинируются лекции, учебные пособия и практические упражнения. Вы научитесь обрабатывать данные о биоразнообразии, включая штрихкоды ДНК. Получите практический опыт использования открытых и задокументированных данных о биоразнообразии через GBIF и BOLD для ответа на вопросы об исследовании биоразнообразия. Вы будете понимать и применять данные наблюдений, коллекционные и генетические данные из аналоговых и цифровых источников. Наконец, этот курс предоставляет базовые навыки публикации данных через GBIF и BOLD.

Сфера применения

Навыки управления данными для получения доступа к данным и их публикации через платформы данных о биоразнообразии. В этом курсе рассматривается наблюдение/экземпляр → публикация записей, не включая лабораторные этапы.

Этот курс представляет собой результат сотрудничества между проектом BioDATA Университета Осло финансируемым Diku, проектом «Кавказский штрихкод жизни» (CaBOL) финансируемым BMBF и GBIF - Глобальным информационным фондом по биоразнообразию, разработан Dag Endresen, Dmitry Schigel, Helena Wirta, Hugo de Boer, Laura Russell, and Stefaniya Kamenova.

Аудитория

Курс ориентирован на аспирантов и кандидатов наук в области биологии, а также других специалистов в соответствующих областях.

Обязательные требования

Участники должны быть связаны либо иметь профессиональный интерес к исследованию биоразнообразия. Участники должны иметь мотивацию и интерес к обработке штрихкодов ДНК, данных музейных коллекций и наблюдений. Хорошее понимание английского языка необходимо, чтобы следовать курсу, выполнять упражнения, и получать поддержку во время обучения.

Предполагаемые результаты обучения

- Понимание и умение объяснять понятие делимитации видов.
- Умение использования данных генетических последовательностей такие как штрихкод ДНК для идентификации вида.
- Умение публикования и извлечения данных из GBIF и BOLD.
- Изучите основы сбора, очистки, хранения, геопривязки и цитирования данных.
- Критически оценивать качество собственных и внешних данных и их пригодность для целей исследования.
- Использование ключевых инструментов и подходов для получения максимально

качественных данных, связывания данных и повторного использования данных.

- Изучите преимущества принципов FAIR и открытых данных в исследованиях биоразнообразия и в сотрудничестве.
- Понимание ценности управления данными как научно-исследовательского инструмента.
- Получение понимания в широком смысле важности международных инфраструктур по биоразнообразию и того, как они могут способствовать его оценке, мониторингу, сохранению и повторному использованию.

Подготовка курса

Необходимо

Регистрация и учетные записи

- Создайте учетную запись в ORCID.
- Создайте личную (пользовательскую) учетную запись на GBIF.org в правом верхнем углу. Вы можете войти в GBIF с помощью ORCID.
- Создайте личную (пользовательскую) учетную запись на boldsystems.org/index.php/MAS_Management_NewUserApp[BOLD^] в правом верхнем углу. Вы можете войти в BOLD с помощью ORCID.
- Создайте учетную запись в iNaturalist.

Основные видео

Чтобы подготовиться к лекциям курса, пожалуйста, ознакомьтесь со списком основных видеоматериалов. Это займет довольно много времени, но будет способствуют тому, чтобы все участники имели одинаковую базу знаний в начале курса и были готовы к лекциям; пожалуйста, записывайте вопросы по мере просмотра материалов и задавайте их в течение курса.

Основные видео - 6 видео, 67 мин.

Рекомендуем

Курс будет включать практическую часть, где вы будете разрабатывать исследовательский проект и его элементы данных. Этот проект и его «путь данных» будут созданы на основе системы растение-опылитель. Для подготовки к этой части, пожалуйста, прочитайте следующее. Пропустите этот раздел, если вы уже работаете с опылителями.

- Детективное исследование меда вызывает опасение за пчел
- Потеря пчел вызывает нехватку основных пищевых культур, результаты исследований

Дополнительно

Следующие действия являются необязательными, если вы выполнили все вышеперечисленное. Многие аспекты из них будут представлены в течение курса.

- Что такое GBIF? - видео, 8 минут
- Что такое BOLD? - видео, 9.5 минут

- Введение в GBIF - онлайн-курс
- Более чем 75% снижение за 27 лет общей биомассы летающих насекомых на охраняемых территориях - статья

Файлы для скачивания

Все файлы курса можно загрузить с этой страницы. Или, если вы предпочитаете иное, все файлы имеют отдельные ссылки на протяжении всего курса, по мере их появления в учебной программе. Видеофайлы также встраиваются в курс и доступны на YouTube. Субтитры доступны при воспроизведении с YouTube для большинства видео. Если у вас есть сложности с доступом к встроенным видео, пожалуйста, загрузите mp4 файлы, чтобы воспроизвести их локально на вашем компьютере.

Видео

Видео представлены на английском языке. Субтитры для загруженных видео недоступны.

Foundations1.zip (73.7 MB)

Foundations2.zip (90.2 MB)

Capture.zip (63.1 MB)

Management.zip (30.2 MB)

Publishing.zip (77.9 MB)

Упражнения

Путь данных, этап 4: sequences.zip файл (ZIP 5 KB)

Путь данных, этап 6:

1. Excel template
2. Vicia.zip file (ZIP 51 MB)

Путь данных, этап 7:

1. ViciaForCleaning.txt file (ZIP 66 KB)
2. UC1-3c-open-refine.csv. (207.5 KB)

Установка программного обеспечения

Установка OpenRefine

Установите программное обеспечение, необходимое для выполнения дальнейших действий в рамках прохождения курса

OpenRefine - это инструмент с набором функций для работы с табличными данными, который улучшает общее качество набора данных. Это приложение работает на вашем компьютере как небольшой веб-сервер, и для того, чтобы использовать его, ваш веб-браузер должен указывать на этот веб-сервер. Итак, подумайте о OpenRefine как о личном и частном веб-приложении.

Мы будем использовать OpenRefine во время похождения раздела курса касающегося мобилизации данных, особенно во время практических занятий. Необходимо установить OpenRefine на ноутбук. Если вы квалифицированный пользователь компьютера, вы можете выполнить следующие действия, чтобы установить программное обеспечение на ваш компьютер. Если вы не уверены, пожалуйста, попросите о помощи. Дополнительные сведения см. на странице загрузки OpenRefine.

Для установки программного обеспечения могут потребоваться пароли администратора.

Требования для установки

1. Только для пользователей Linux: установлен Java JRE.
2. Google Chrome, Microsoft Edge или Mozilla Firefox установлены. Internet Explorer не поддерживается.

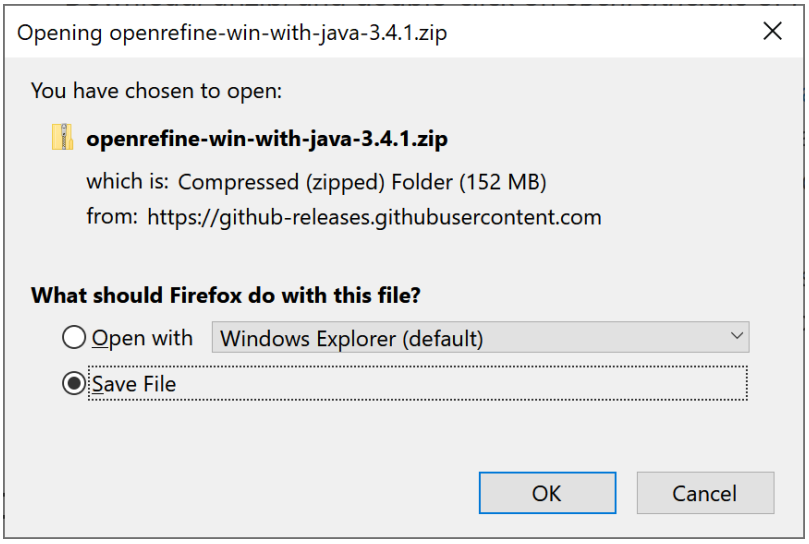
Последняя стабильная версия OpenRefine 3.4.1, выпущенная 24 сентября 2020 года. Подробные инструкции по установке можно найти на https://docs.openrefine.org/manual/installing/.

Установка на Windows

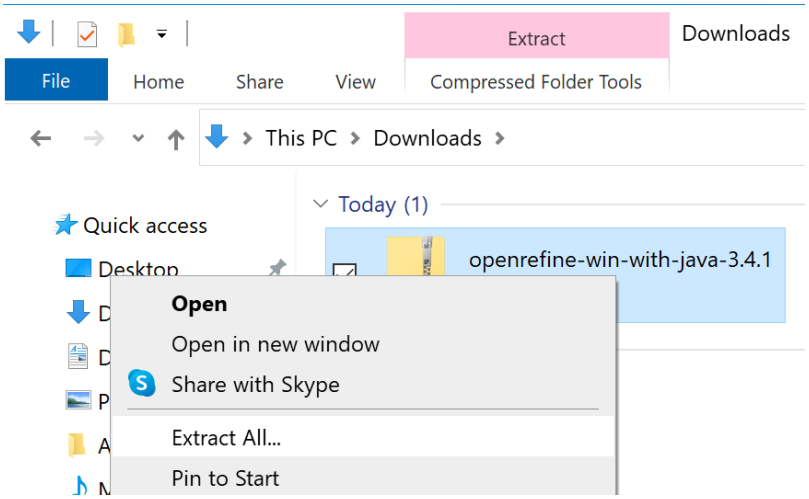
1. Скачайте Windows kit **with embedded Java** (Windows со встроенной Java). Выберите сохранить файл, а не открыть его.
2. Найдите скачанный файл. Щелкните правой кнопкой мыши и выберите "Extract all..." (Извлечь все...). Разархивируйте и дважды щелкните на openrefine.exe или refine.bat, если первая не работает.
3. Появится командное окно (не закрывайте его), и вскоре после этого новое окно веб-браузера покажет приложение.

▼ *Подробные инструкции для MS Windows (нажмите, чтобы развернуть)*

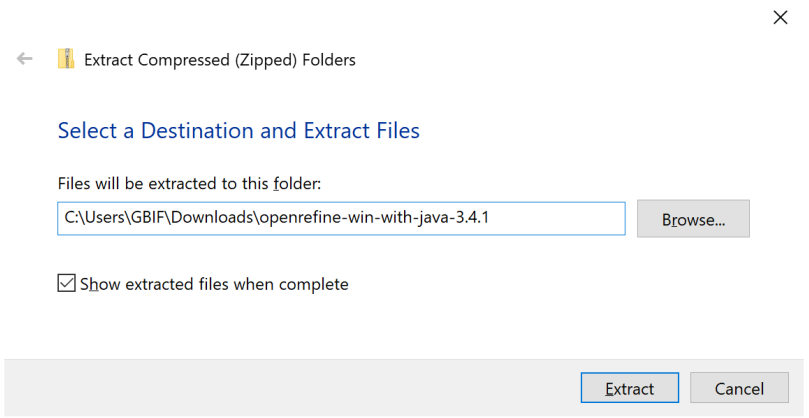
1.



Скачайте Windows kit **with embedded Java** (Windows со встроенной Java). Выберите сохранить файл, а не открыть его.

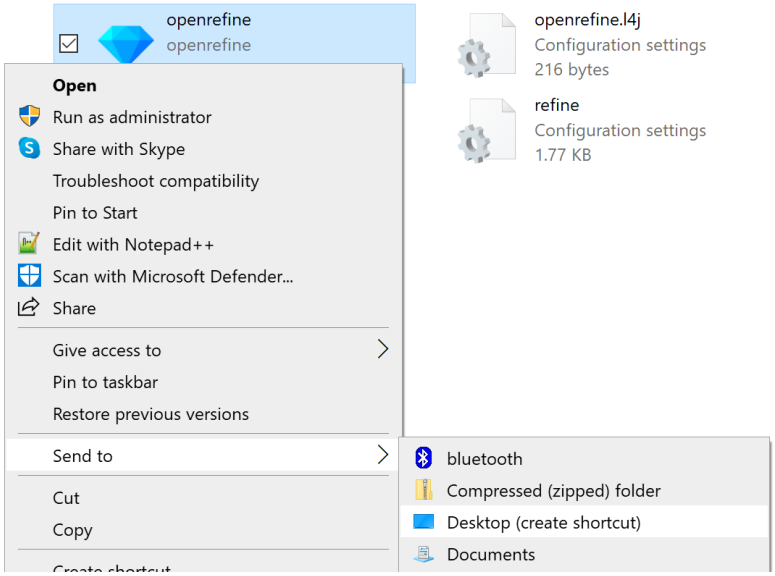
2. 

Найдите файл, который вы скачали. Щелкните правой кнопкой мыши и выберите "Extract All..." (Извлечь все...)

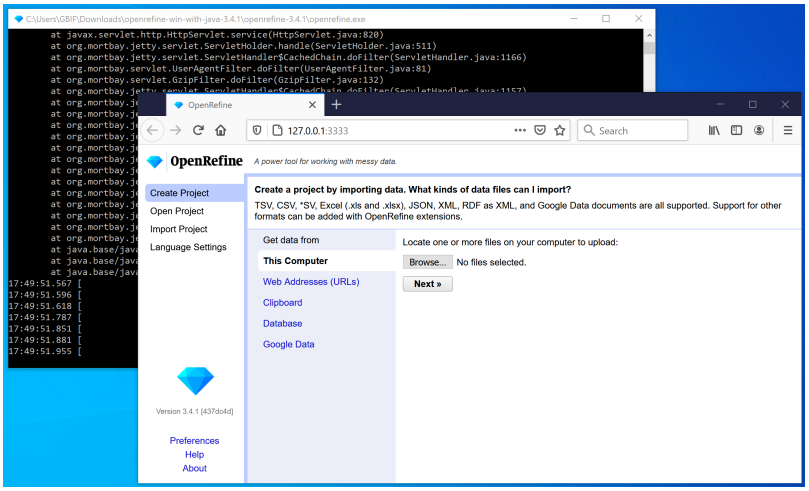
3. 

Нажмите "Extract" (Извлечь)

4.



Найдите извлеченные файлы. При необходимости щелкните правой кнопкой мыши "openrefine" и выберите "Send to → Desktop (create shortcut)", чтобы создать ярлык на рабочем столе. Затем дважды нажмите "openrefine"

5. 

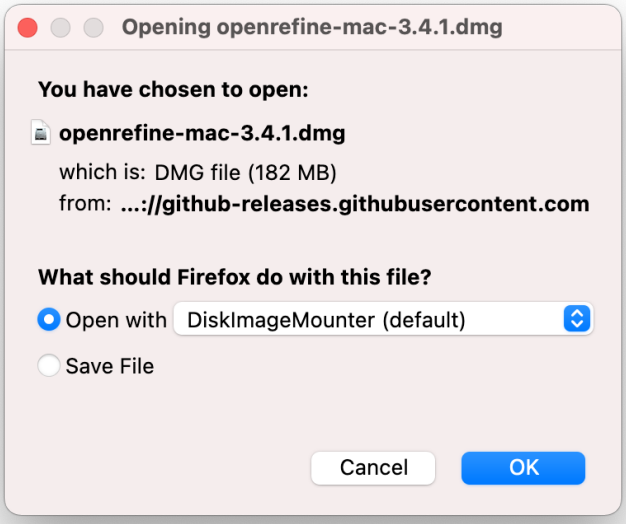
Откроется черное окно консоли, а через некоторое время откроется браузер. OpenRefine теперь готов к использованию.

Установка на Mac

1. Скачайте Mac kit.
2. Загрузите, откройте, перетащите значок в папку Приложения. Отдельно устанавливать Java не требуется.
3. Дважды щелкните по нему, и новое окно веб-браузера покажет приложение.

▼ *Подробные инструкции для Mac (нажмите, чтобы развернуть)*

1.



Скачайте Mac kit и откройте его.

2.

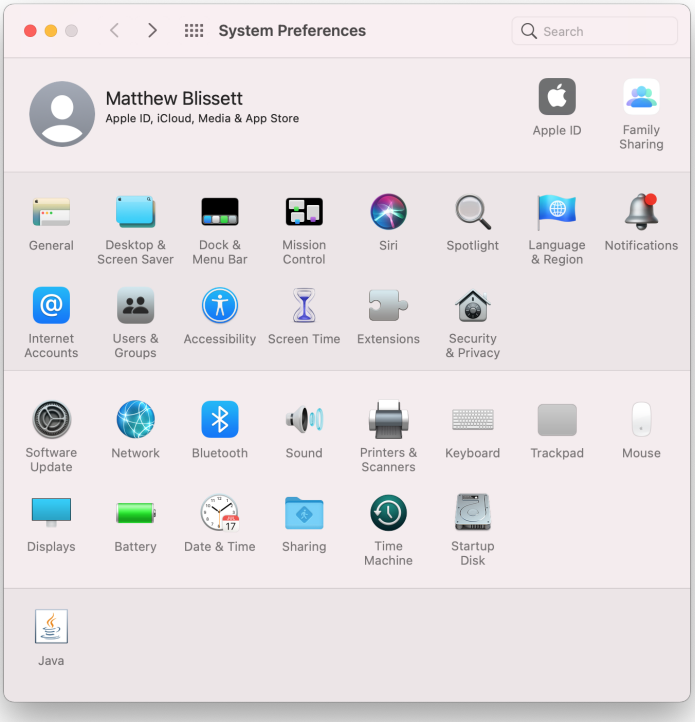
Показано предупреждение. Нажмите "OK".

3.

System Preferences

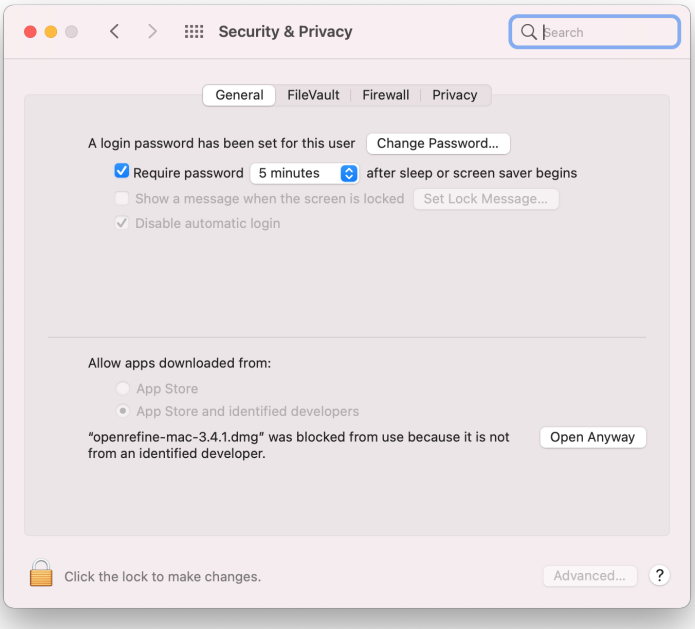
Откройте Системные настройки.

4.



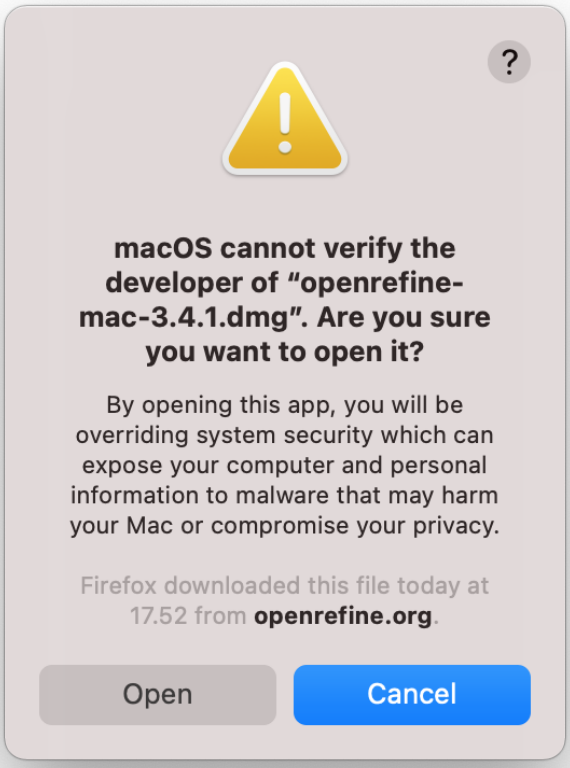
Безопасность и конфиденциальность

5.



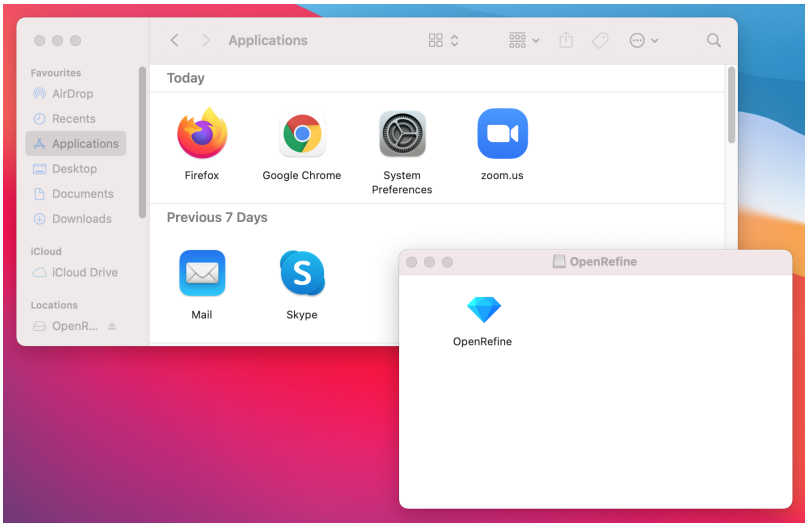
В нижней части выберите "Open Anyway" (Открыть в любом случае).

6.



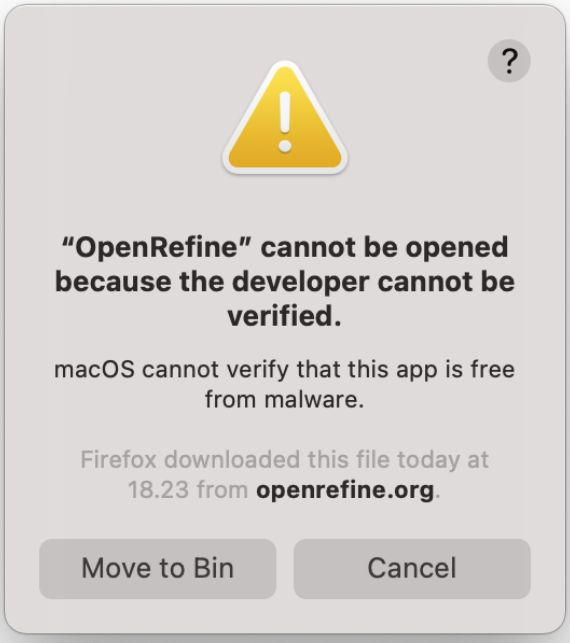
Выберите "Open" (Открыть)

7.



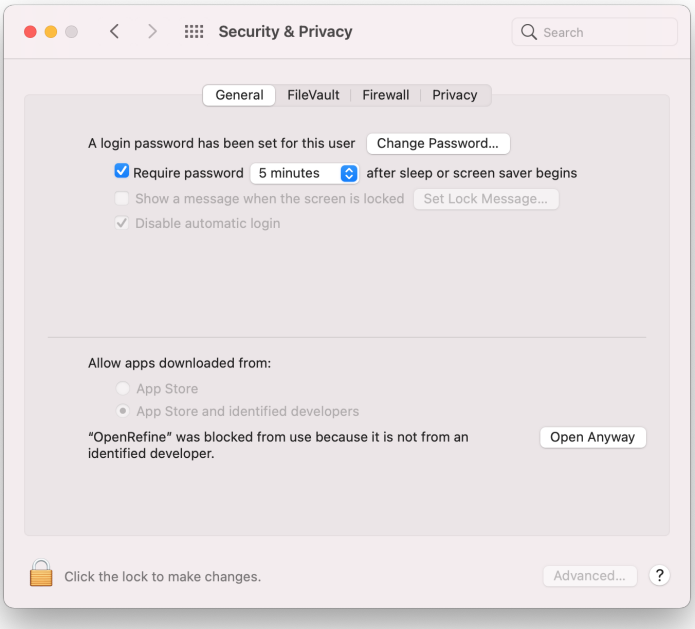
В конечном итоге открывается архив приложения! Перетащите его в папку Приложения.

8.



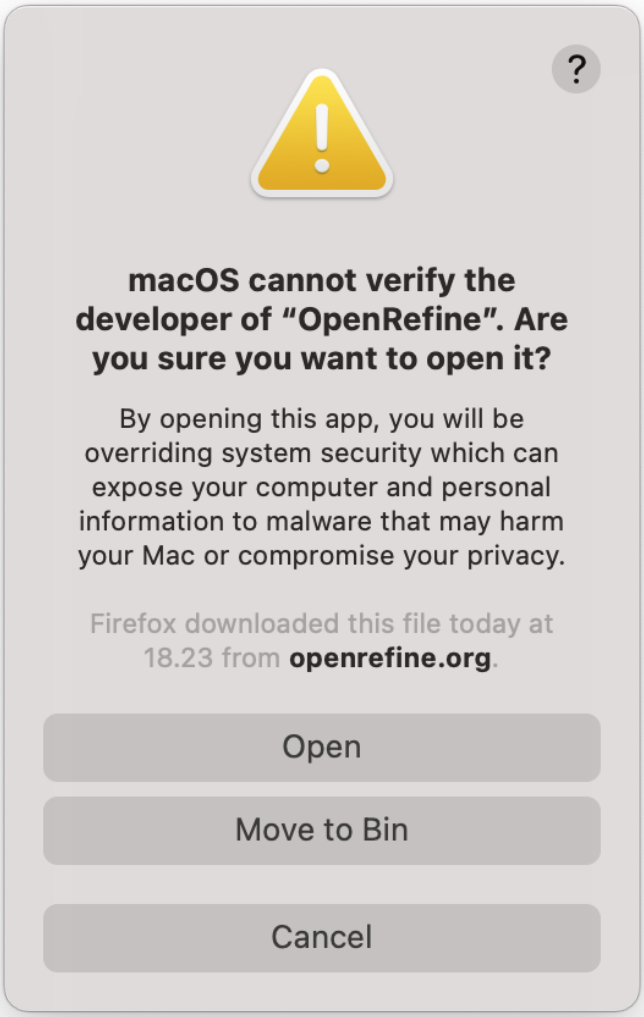
Дважды щелкните значок OpenRefine. Появится еще одно предупреждение безопасности!

9.



Вернитесь в "Безопасность и конфиденциальность" и снова нажмите "Open Anyway" (Открыть в любом случае).

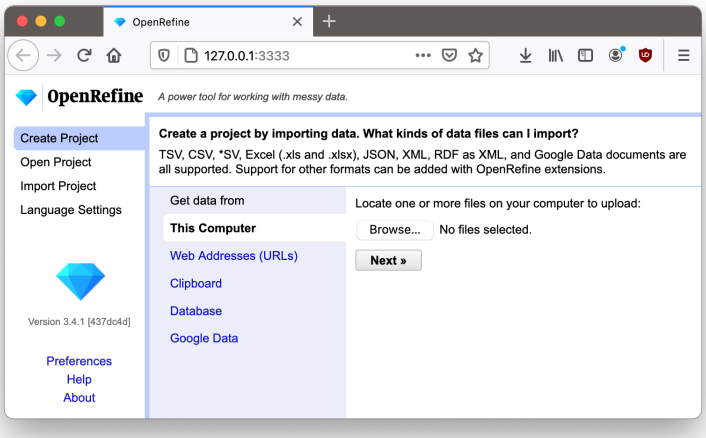
10.



(Чтобы избежать этих предупреждений, разработчикам OpenRefine придется заплатить Apple.)

Нажмите "Открыть".

11.



Наконец! Приложение запущено.

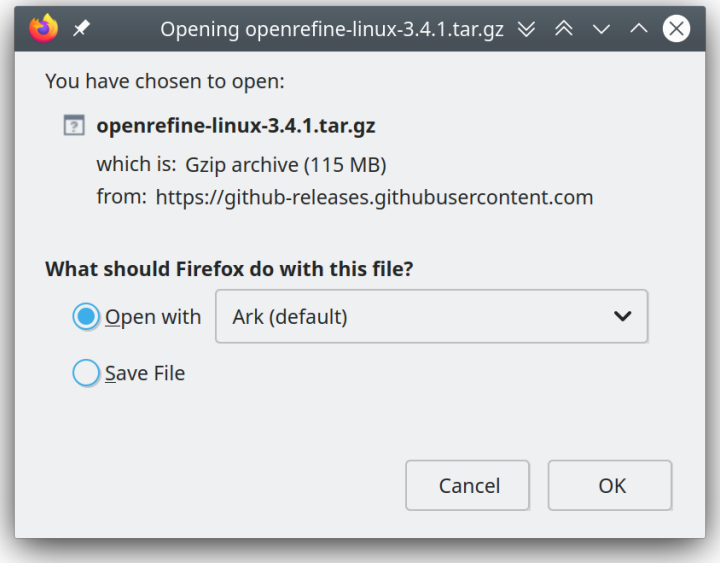
Установка на Linux

1. Скачайте Linux kit.
2. Загрузите, извлеките, затем введите ./refine для запуска. Потребуется установка Java на вашем компьютере.

▼ *Подробные инструкции для Linux (нажмите для расширения)*

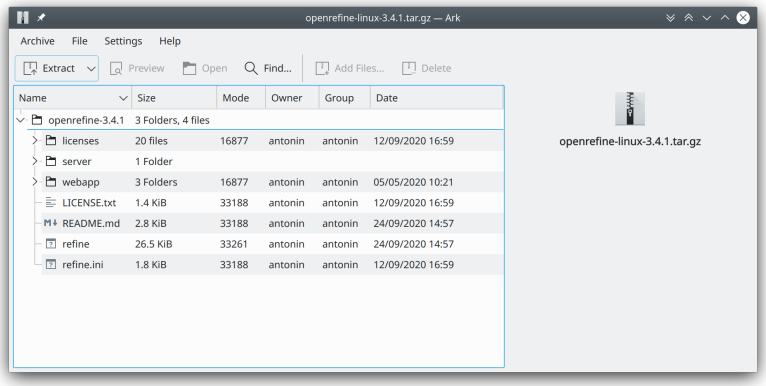
Эти инструкции предназначены для KDE (например, Kubuntu, SuSE), но процесс похож на Gnome (например, Ubuntu, Red Hat, CentOS).

1.



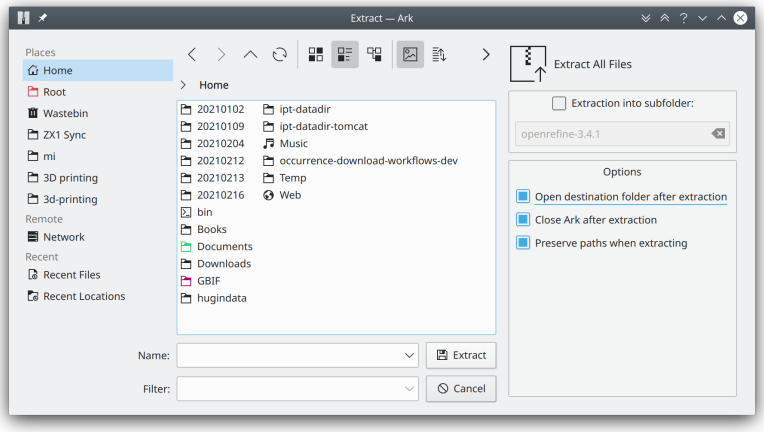
Скачайте Linux kit.

2.

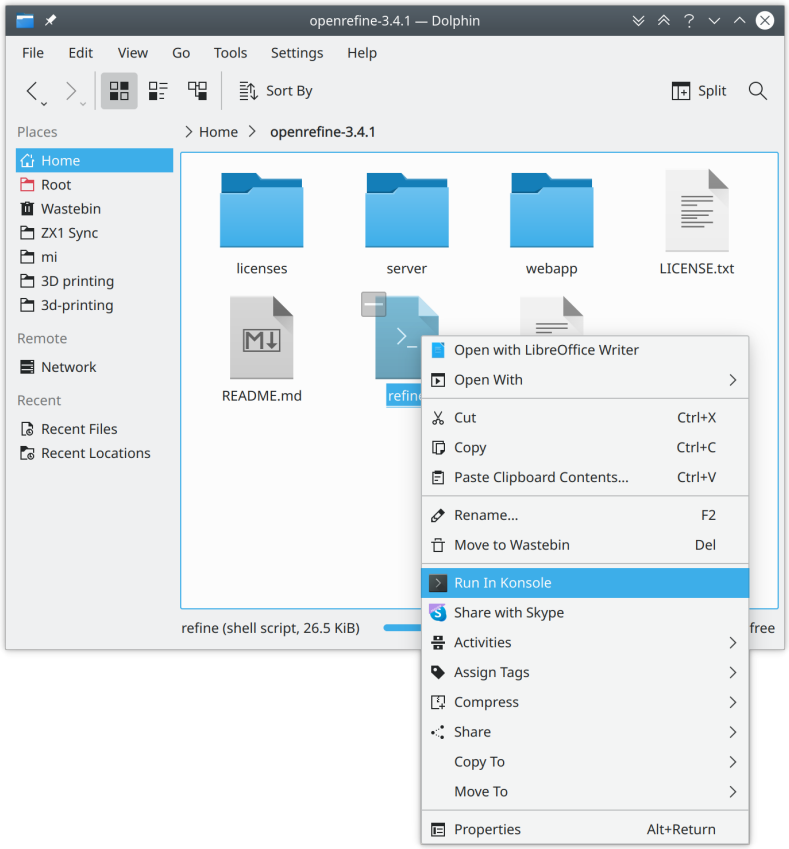


Нажмите "Extract" (Извлечь) для распаковки загруженного приложения.

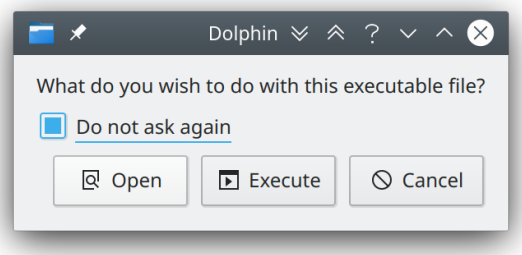
3.



Выберите подходящее место. Например, «Open destination folder after extraction» (Открыть папку назначения после извлечения) и «Close Ark after extraction» (Закрыть архив после извлечения)

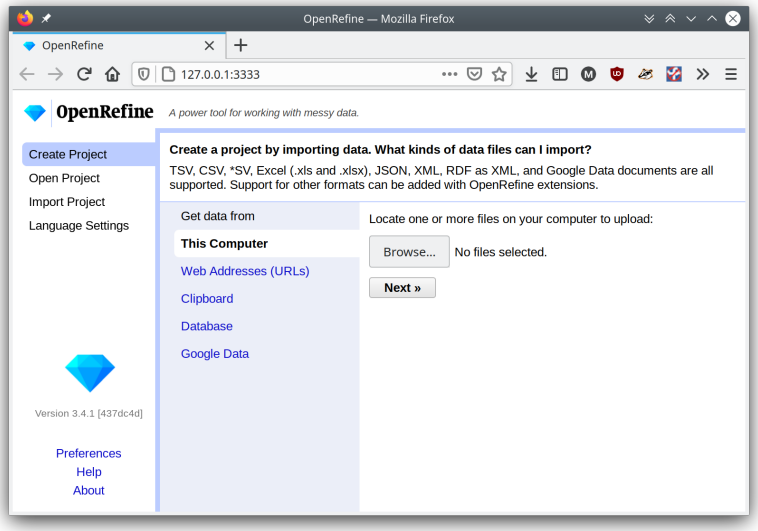
4. 

Щелкните правой кнопкой мыши "refine" и выберите "Run in Konsole". Это необходимо для безопасного выхода из OpenRefine позже, закрыв окно Konsole.

5. 

Подтвердите, что вы хотите выполнить загруженное приложение.

6.



OpenRefine теперь запущен.

1. Введение

1.1. Biodiversity of Bulgaria

Презентация доступна в онлайн-версии курса.

1.2. Штрихкод, данные и биоразнообразие

Презентация доступна в онлайн-версии курса.

1.3. GBIF для науки и политики

Презентация доступна в онлайн-версии курса.

2. Данные как первоклассные исследовательские объекты

Этот модуль включает информацию о тенденциях для рассмотрения исследовательских данных как первоклассных научных объектах, для независимого цитирования и внесения вклада в оценку карьеры исследователя.

2.1. Открытые данные исследований

В этой презентации вы рассмотрите последние тенденции относящиеся к исследовательским данным как первоклассных научным объектам, которые будут независимо цитироваться, и вносить вклад в оценку карьеры исследователя, используемую в этом курсе.

Презентация доступна в онлайн-версии курса.

3. Путь данных - шмели опылители

Путь данных представляет собой серию практических упражнений, посвященных шмелям опылителям. Используя GBIF и BOLD, вы узнаете: I. как найти и использовать уже имеющиеся данные о биоразнообразии в связи с вашими исследовательскими вопросами; II. Как эффективно фиксировать и очищать эти данные, т.е. перевести их в стандартный формат, который имеет прямое отношение к вам и пригоден для вашего использования; III. Как генерировать и публиковать новые данные в соответствии с международными стандартами.

Перемещение по данным состоит из девяти этапов. Каждый этап (или серия этапов) соотносится с различными модулями курса. После серии теоретических лекций вы вернетесь к Пути данных, чтобы завершить практические занятия. Практические занятия идут по следующему пути: I. система обучения; I. вопросы и гипотезы; III. доступность данных; IV. сбор и очистка данных; и V. создание и публикация данных.

Презентация доступна в онлайн-версии курса.

4. Идентификация вида

4.1. Зачем нам нужно определить виды

Презентация доступна в онлайн-версии курса.

4.1.1. Путь данных, этап 2

Завершите второй этап, задачи 4-6

4.1.2. Путь данных, этап 3

Завершите этап 3, задачи 7-9

4.2. Идентификация и делимитация видов ДНК

Презентация доступна в онлайн-версии курса.

5. Депозитарии штрихкодов

5.1. Депозитарии штрихкодов

Презентация доступна в онлайн-версии курса.

5.2. Демонстрация

5.3. Путь данных, этап 4

Завершите этап 4, задача 10.

5.4. Путь данных, этап 5

Завершите этап 5, задача 11.

Unresolved directive in index.ru.adoc - include::scholarly-recognition.ru.adoc[]

Unresolved directive in index.ru.adoc - include::unite.ru.adoc[]

6. Приложения для меташтрих-кодирования и eDNA

Презентация доступна в онлайн-версии курса.

Unresolved directive in index.ru.adoc - include::bold.ru.adoc[]

Unresolved directive in index.ru.adoc - include::iNaturalist.ru.adoc[]

7. Сбор данных

В этом модуле вы узнаете типы первичных данных о биоразнообразии и как наилучшим образом обмениваться этой информацией в рамках GBIF. Вы также рассмотрите принципы качества данных в контексте сбора данных и узнаете о качестве и согласованности данных (особенно по таким вопросам, как географическое распространение, даты, названия и перекрестная таксономическая проверка).

7.1. Источники и типы данных

В этом видеоролике (10:45) вы рассмотрите https://www.gbif.org/dataset-classes [первичные данные о биоразнообразии ^], которыми можно поделиться в рамках GBIF. Если вы не можете посмотреть встроенное видео, вы можете скачать его по ссылке download. (MP4 - 19 МБ)

▶ https://www.youtube.com/watch?v=wKeOveydjsw *(YouTube video)*

7.1.1. Обзор типов данных

Проверьте себя на освоение понятий, изученных в этом разделе.

1. Какие типы или тип наборов данных вы выберите для ихтиологической коллекции?

   - ☐ occurrence (наблюдение)
   - ☐ checklist (таксономический список)
   - ☐ sampling event (отбор проб)

2. Какие типы или тип наборов данных вы выберите для списка инвазивных видов?

   Водяной гиацинт (Eichhornia crassipes), наблюдаемая в Бурае, Новая Каледония, где он является интродуцированным инвазивным видом согласно GRIIS. Фото gerard (2016) лицензировано под CC BY-SA 2.0 изображением::img/web/QDataTypes-plant.png[align=center,width=640,height=360]

   - ☐ occurrence (наблюдение)
   - ☐ checklist (таксономический список)
   - ☐ sampling event (отбор проб)

3. Какие типы или тип наборов данных вы выберите для исследования воздействия на окружающую среду?

   Исследования по оценке воздействия на окружающую среду проводятся экспертами для определения биоразнообразия и биотопов в данном районе до, во время и после воздействия на них со стороны деятельности человека (дорожные работы, ветровые турбины, добыча полезных ископаемых, строительство зданий и т. д.).

Энтомолог ловит бабочек *фото Matthieu Gauvain (CC-BY-SA)*

- ☐ occurrence (наблюдение)
- ☐ checklist (таксономический список)
- ☐ sampling event (отбор проб)

4. Какие типы или тип наборов данных вы выберите для данных отслеживания перемещений птиц?

   Данные отслеживания перемещений птиц регистрируются с использованием специальных устройств, например GPS-трекеров, установленных на живых птицах, это позволяет ученым отслеживать их маршруты миграции или участки гнездования.

   Griffin vulture *наблюдалась в Заповеднике Гамлы фото גיזונים - MinoZig (CC0)*

   изображение::img/web/QDataTypes-tracking.png[align=center,width=640,height=360]

   - ☐ occurrence (наблюдение)
   - ☐ checklist (таксономический список)
   - ☐ sampling event (отбор проб)

5. Какие типы или тип наборов данных вы выберите для данных по ловушкам насекомых?

Ловушка насекомых *фото miheco (CC-BY-SA)*

- ☐ occurrence (наблюдение)
- ☐ checklist (таксономический список)
- ☐ sampling event (отбор проб)

6. Какие типы или тип наборов данных вы выберите для данных управления национальными парками?

   Данные, полученные в контексте управления охраняемыми районами (такие, как национальные парки, а также и более мелкие природные охраняемые территории), могут быть разнообразными и иметь различное происхождение: ботанические обследования, слежение за животными, наблюдения со стороны рейнджеров и охранников, а также данные от "гражданской науки", полученные с фотографий, размещенных на социальных сетях.

   слоны Шри-Ланки *наблюдались pen_ash.*

   изображение::img/web/QDataTypes-Observations.png[align=center,width=640,height=360]

   - ☐ occurrence (наблюдение)
   - ☐ checklist (таксономический список)
   - ☐ sampling event (отбор проб)

7. Какие типы или тип наборов данных вы выберите для биоблица гражданской науки?

   Данные гражданской науки зачастую собираются во время тематических полевых дней, известных как «биоблиц». Волонтеры, как правило, собираются в конкретном регионе и проводят день наблюдая и определяя столько видов, сколько смогут найти на его территории.

   Данные каждого участника фиксируются и объединены в программу по сбору данных

или управлению данными.

Поиск птиц с персоналом парка *фото Национальной парковой службы США (авторизованное повторное использование в поиске изображений Google)*

- ☐ occurrence (наблюдение)
- ☐ checklist (таксономический список)
- ☐ sampling event (отбор проб)

8. Какие типы или тип наборов данных вы выберите для регионального списка видов?

Black rhino *наблюдался в зоопарке Магдебурга в Германии, фото Mani300*

- ☐ occurrence (наблюдение)
- ☐ checklist (таксономический список)
- ☐ sampling event (отбор проб)

7.1.2. Анализ решений по типам данных

▼ *Нажмите, чтобы раскрыть*

Какие типы или тип наборов данных вы выберите для ихтиологической коллекции?

- occurrence
  Most of the time, specimens from collection databases are shared as occurrence data. Each occurrence (specimen or group of specimens) has its own unique identifier (sometimes derived from its catalogue number in the source collection) and the Darwin Core fields used to share them within GBIF describe each specimen: scientific name, the date it was collected on the field, who collected and/or identified it, where, etc. Each collection can have more than one specimen from a same species, as long as each specimen is identified by a unique ID.
- checklist
  It is also possible to create and share a taxonomical checklist derived from a collection database; in this case, it is recommended to share the checklist as a taxonomical dataset, with the occurrence (specimen) list associated with it by using the Occurrence core as an extension to the Taxon Core on the GBIF IPT.

Какие типы или тип наборов данных вы выберите для списка инвазивных видов?

- occurrence
  Some data publishers will share occurrence datasets coming from studies or programs tracking specimens from some specific invasive species; when the data focuses on individuals instead of the invasive species, in general, they can be shared as occurrence data.
- checklist

Invasive species can be tracked and monitored at different scales (regional, national, thematic…); as this type of dataset focuses more on the species and their distribution across a given geographical scope, they are mainly shared as taxonomical datasets within GBIF (see GRIIS search results).

Какие типы или тип наборов данных вы выберите для исследования воздействия на окружающую среду?

- occurrence
  Data are recorded by naturalists on the field and can be shared as simple occurrence datasets.
- sampling event
  They can also be shared as event datasets if standardized protocols (such as vegetation plots, transects, traps…) are used to collect the data.

Какие типы или тип наборов данных вы выберите для данных отслеживания перемещений птиц?

- occurrence
  These data are shared as occurrence datasets: ideally, each bird is identified with its organismID, and each occurrence (GPS ping) has its own occurrenceID, which is useful to track the different GPS locations of the same bird over the scope of the tracking programme or project. (See example)

Какие типы или тип наборов данных вы выберите для данных по ловушкам насекомых?

- occurrence
  Although such data can be shared as simple occurrence datasets, it is best if they're shared as event datasets, where the location, identifier and contents of each trap can be better detailed.
- sampling event
  Insect traps (as well as other traps such as pitfall traps, malaise traps…) are typically used in monitoring programmes to check the presence (or absence) of some species and/or assess their specific abundance. Using the "eventID" field to identify each trap allows the users to get all of the specimens collected within each trap. The same logic applies to other field protocols such as transects, plots, remote cameras, etc.: by using the Event Core instead of the Occurrence core, you'll be able to share much more information about the context of the data collection, and allow users to better understand (and even replicate) your work.

Какие типы или тип наборов данных вы выберите для данных управления национальным парком?

- occurrence
  record individuals of species
- checklist
  It is important to know how many species are present in the park/reserve perimeter and their conservation status.
- sampling event
  check and track the populations

Какие типы или тип наборов данных вы выберите для биоблица гражданской науки?

- occurrence
  Bioblitz datasets are mainly shared as occurrence datasets.
- sampling event

Depending on the citizen science programme, specific sampling protocols might be used by the volunteers, in which case, the data can be shared as an event dataset.

Какие типы или тип наборов данных вы выберите для регионального списка видов?

- checklist
  Geographical or thematic species lists are often used to share information about the species present in a given area; most of the time, these lists also mention the distribution of each species as well as their conservation status in this area. Regional species lists can give a useful insight into a region's biodiversity and habitats, and need to be shared as taxonomical datasets, with or without associated occurrences.

7.2. Сбор данных, их обработка и качество

В этом видео (09:11) вы изучите основы качества данных, относящиеся к их сбору, применительно к извлечению данных из коллекционных этикеток, полевых дневников, электронных таблиц и т.д. Если вы не можете смотреть встроенное видео, вы можете использовать ссылку для его просмотра удаленно download. (MP4 - 19 MB)

▶ https://www.youtube.com/watch?v=QkDJIkmwBMA *(YouTube video)*

7.3. Путь данных, этап 6

Завершите этап 6, задание 12.

8. Управление данными

В этом модуле вы рассмотрите основные концепции, связанные с этим инструменты и лучшие примеры существующих методов управления данными, в частности, очистки данных и их стандартизации.

8.1. Принципы управления данными

В этом видео (09:49) вы узнаете о различных инструментах, которые вы можете использовать для улучшения качества ваших данных. Если вы не можете смотреть встроенное видео, вы можете воспользоваться ссылкой для его просмотра удаленно download. (MP4 - 16.6 MB)

▶ https://www.youtube.com/watch?v=4ijm1cJeVHE *(YouTube video)*

8.2. Инструменты управления данными

В этом видео (06:42) вы узнаете о различных инструментах, которые вы можете использовать для улучшения качества ваших данных. Если вы не можете смотреть встроенное видео, вы можете воспользоваться ссылкой для его просмотра удаленно download. (MP4 - 10.3 MB)

▶ https://www.youtube.com/watch?v=Ru3vWiYU3gw *(YouTube video)*

8.3. OpenRefine

В этом видео (03:27) вы узнаете о OpenRefine. Вы можете использовать OpenRefine для стандартизации и улучшения качества ваших данных. Если вы не можете смотреть встроенное видео, вы можете воспользоваться ссылкой для его просмотра удаленно download. (MP4 - 3.8 MB)

▶ https://www.youtube.com/watch?v=_YFw_bfwc3Y *(YouTube video)*

8.4. Путь данных, этап 7

Завершите этап 7, задания 13-15.

8.5. Список упражнений

8.5.1. Проверка валидации

Technical errors (Технические ошибки) Относительно простой, часто поддающийся автоматизации, **проверяет целостность данных**. Они могут указывать на некорректный экспорт, отображение данных, смещение поля (например, перемещение 1 колонки справа) или отсутствие данных в источнике.

- **Completeness** (Укомплектованность): Доступны ли все данные и метаданные – присутствуют ли все поля, заполнены ли все поля?
- **Bounds** (Пределы): Например, дни указаны в диапазоне 1-31 (в зависимости от месяца)
- **Data type** (Тип данных): Например, поле "Дата" содержит дату или цифру?
- **Data format** (Формат данных): Например, даты представлены 01/01/2010 или 01/Ян/10?

Consistency errors (Ошибки соответствия)

Применение правил реального мира к данным. Они могут указывать на некорректный ввод данных из старых записей, ошибок транскрипции или последующей обработки. Некоторые из них сложны для реализации и **требуют справочных наборов данных для повторной проверки**. Например, список известных коллекторов и способов сбора коллекций. Эти правила могут быть собраны у пользователей и аналитиков.

- **Taxonomic** (Таксономия): Например, в случае идентификации видового уровня, имеются ли бинарное научное название и записи в полях родов и видов?
- **Currency** (Обмен): Согласованы ли даты сбора, идентификации, обновления и оцифровки?
- **Outliers** (Отклонения): Находите отклонения, но помните, что не все отклонения обязательно являются ошибками. Например, сравните с известным спектром видов или известным экологическим диапазоном (но помните, что отклонения могут быть неправильной идентификацией, а не неправильными координатами).
- **Geographic** (География): Координаты в пределах идентифицированного населенного пункта или региона? Например, имеются ли какие-либо наземные наблюдения на море или морские наблюдения на суше?
- **Collecting patterns** (Модели коллектирования): Соответствует ли подробная информация о наблюдении известным моделям коллектирования организации или сборщика? Возможно ли создание каких-либо записей после смерти сборщика (возможно, это другой сборщик с

похожим именем)? Например, приписываются ли какие-либо записи о млекопитающих группе, наблюдающей за птицами?

- **Accuracy and precision** (Аккуратность и точность): Например, являются ли какие-либо географически привязанные записи, указывающие на очень высокую точность или аккуратность сбора данных в период до появления GPS (или до точности GPS)?
- **Collecting methods** (Методы сбора): Различные методы обследования (например, трансекты и обследования целых районов) имеют особые характеристики. Согласуются ли записи с предоставленным методом?

8.5.2. Полезные инструменты

- **GBIF Name Parser**: https://www.gbif.org/tools/name-parser
- **Global Names Resolver**: http://resolver.globalnames.org
- **Catalogue of Life name match**: https://data.catalogueoflife.org/tools/name-match
- **TNRS**: https://tnrs.biendata.org/
- **WoRMS**: https://www.marinespecies.org/aphia.php?p=match
- **InfoXY**: http://splink.cria.org.br/infoxy?criaLANG=en
- **Georeferencing Calculator**: http://georeferencing.org/georefcalculator/gc.html
- **Canadensys coordinate conversion**: http://data.canadensys.net/tools/координаты
- **Canadensys date parsing**: http://data.canadensys.net/tools/dates
- **Карты Google**: https://maps.google.com/

9. Публикация данных

В этом модуле вы узнаете об основах публикации данных, включая IPT, ядра и расширения, и важности лицензий, метаданных, обязательных полей и размещения наборов данных.

9.1. Концепции публикации данных

В этом видео (11:45) вы узнаете об основах публикации данных и получите введение в Integrated Publishing Toolkit (https://www.gbif. Если вы не можете посмотреть встроенное видео, вы можете воспользоваться ссылкой для просмотра видео удаленно download. (MP4 - 20 MB)

▶ https://www.youtube.com/watch?v=b900d9ukjSQ *(YouTube video)*

9.2. Обзор IPT

В этом видео (06:56) вы получите обзор интерфейса публикации данных IPT. Если вы не можете смотреть встроенное видео, вы можете воспользоваться ссылкой для его просмотра удаленно download. (MP4 - 8.7 MB)

▶ https://www.youtube.com/watch?v=gHXsaN_JWeI *(YouTube video)*

9.2.1. Обучение установке IPT

If you have interest in trying the IPT, please contact training@gbif.org and you will be provided with a login and password to one of the training IPTs.

https://training-ipt-a.gbif.org/

https://training-ipt-b.gbif.org/

https://training-ipt-c.gbif.org/

9.3. Демонстрация IPT

В этом видео (24:16) вы узнаете, как опубликовать набор данных с помощью IPT. Если вы не можете смотреть встроенное видео, вы можете воспользоваться ссылкой для его просмотра удаленно download. (MP4 - 52.6 MB)

▶ https://www.youtube.com/watch?v=eDH9IoTrMVE *(YouTube video)*

9.4. Путь данных, этап 8

Complete step 8, task 16 using https://cloud.gbif.org/eca

10. Заключение

Подготовка курса

Завершите оценку курса

Оценка может быть завершена по ссылке online.

Словарь

ALA

Atlas of Living Australia (Атлас живой природы Австралии). Австралийский узел GBIF, который разработал портал данных с открытым исходным кодом, в настоящее время широко используется в рамках сообщества и партнеров GBIF для национальных порталов биоразнообразия.

API

Application Programming Interface (Интерфейс прикладного программирования). Набор четко определенных способов связи между различными программными компонентами.

BID

Biodiversity Information for Development (Информация о биоразнообразии в целях развития). Финансируемый ЕС проект, координируемый GBIF и направленный на расширение возможностей по мобилизации данных в регионах Африки, Карибского бассейна и Тихого океана.

BIFA

Фонд биоразнообразия Азии.

Лицензии CC

Creative Commons. Это серия лицензий, созданных организацией Creative Commons, которые позволяют обмениваться и повторно использовать творчество и знания посредством предоставления бесплатных юридических инструментов. Три из них могут быть назначены совместно используемым наборам данных GBIF: CC0, CC BY и CC BY-NC.

[Контролируемый словарь

Это ограниченный набор терминов, которые используются в качестве возможных значений для данного поля. Его можно рассматривать как список подстановки или раскрывающийся список для определенного поля. Например, область DwC basisOfRecord должно содержать только одно из этих значений: «PreservedSpecimen», «FossilSpecimen», «LivingSpecimen», «HumanObservation», «MachineObservation». Мы бы сказали, что список значений является контролируемым словарем для этой области.

DwC

Darwin Core - это стандарт данных о биоразнообразии, поддерживаемый TDWG и широко используемый в сообществе и партнерах GBIF. Это набор стандартизированных терминов (или названий полей) и их определений, которые используются для обмена информацией о биоразнообразии.

DOI

Digital Object Identifier (Цифровой идентификатор объекта). Постоянный идентификатор или дескриптор, используемый для уникальной идентификации объектов. DOI широко используются в основном для выявления академической, профессиональной и правительственной информации, такой как журнальные статьи, исследовательские отчеты и наборы данных, а также официальных публикаций.

DwC-A

Архив Darwin Core. Сжатый (архивированный) файл, содержащий всю информацию, необходимую для совместного использования в GBIF для определенного ресурса. Каждый zip содержит три типа файлов:

1. фактические данные в одном или нескольких текстовых файлах: occurrence.txt/event.txt/measurmentoffact.txt и т.д
2. файл сопоставления: rtf.xml
3. файл метаданных (EML) eml.xml. При публикации с помощью IPT создается архив Darwin Core, совместно используемый в GBIF. Кроме того, при загрузке данных с сайта GBIF можно также выбрать формат DwC-A.

GUID

Globally Unique Identifier (Глобальный идентификатор уникальных имен)

IPT

Integrated Publishing Toolkit (Интегрированный набор средств публикации). Это бесплатное и открытое веб-приложение (программное обеспечение) для публикации данных о биоразнообразии. Само программное обеспечение находится на сервере (либо в вашем учреждении, либо в другом месте), который должен иметь доступ к интернету 24/7. Он используется для создания и обработки файлов Darwin Core Archive, которые могут совместно использоваться любым пользователем, включая GBIF.

Заимствование

В контексте коллекций естественной истории это процедура заимствования образцов между учреждениями.

LSID

Life Sciences Identifier (Естественно-научный идентификатор). Они являются постоянными, уникальными в глобальном масштабе идентификаторами биологических объектов.

[Публикация данных

По отношению к GBIF, у нас есть очень конкретное определение публикации данных. Речь идет о том, чтобы сделать наборы данных о биоразнообразии общедоступными и обнаруживаемыми в стандартизированной форме через точку доступа, как правило, через веб-адрес (URL).

Ресурсы

Ресурс - это собирательный термин, используемый для ссылки на определенный набор данных и его метаданные после загрузки на конкретный IPT.

TDWG

Taxonomic Databases Working Group (Рабочая группа по таксономическим базам данных), которая в настоящее время переименована в Biodiversity Information Standards (Стандарты информации о биоразнообразии).

URN

Uniform Resource Number (Унифицированный номер ресурса)

UUID

Универсальный уникальный идентификатор

Колофон

Предлагаемая цитата

Ускорение исследований биоразнообразия с помощью ДНК-баркодов, сбора и наблюдения данных. 1-е издание. Секретариат GBIF: Копенгаген. https://doi.org/10.35035/DZE9-MP34. [Дата проведения курса]

Участники

Этот курс представляет собой результат сотрудничества между проектом BioDATA Университета Осло финансируемым Diku, проектом «Кавказский штрихкод жизни» (CaBOL) финансируемым BMBF и GBIF - Глобальным информационным фондом по биоразнообразию, разработан Dag Endresen, Dmitry Schigel, Helena Wirta, Hugo de Boer, Laura Russell и Stefaniya Kamenova.

Благодарности

Иконки, созданные Freepik с flaticon.com.

Лицензия

Постоянный URI

https://doi.org/10.35035/DZE9-MP34

Управление документами

Первое издание, июнь 2021 года