

# Навыки управления данными в образовании и исследованиях биоразнообразия Беларуси

Повышение квалификации в мобилизации данных о биоразнообразии

11–15 ноября 2019 года, Беловежская пуца, Беларусь

*Последнее обновление 2019-11-07*



Фото: [европейский олень \(Cervus elaphus\) Беловежская пуца март 2017, CC-BY Александр Болбот \(iNaturalist\)](#).

## Введение

[Учебный курс](#) направлен на повышение умений эффективно планировать и реализовывать усилия по оцифровке и мобилизации данных о биоразнообразии в соответствии со стандартами GBIF. Большое внимание будет уделено техническим аспектам мобилизации данных, в частности, всему, что связано с жизненным циклом данных: планированию, сбору данных, их качеству и публикации. Курс нацелен на увеличение объема, полноты и качества данных, публикуемых через сеть GBIF, доступных для использования в научных исследованиях и принятия решений в области природопользования. Также будут рассматриваться социальные аспекты этого процесса. Учебный курс будет состоять из онлайн и очного компонентов, которые будут иметь практический характер, включая значительный объем групповой работы.

Курс будет проходить с 11 по 15 ноября 2019 г. в Беларуси. Он организован Национальной академией наук Беларуси, Музеем естественной истории Университета Осло, норвежским “узлом” GBIF, ForBio и Секретариатом Глобальной информационной системы по биоразнообразию (GBIF). Курс является частью проекта [BioDATA](#) (2018–2021), финансируемого *Норвежским агентством международного сотрудничества и повышения качества высшего образования (DIKU)*.

## Тематики курса

### Основы информатики биоразнообразия

- Ключевые концепции информатики биоразнообразия, в том числе, управление цифровыми данными о биоразнообразии.
- Введение в стандарт Darwin Core и его компоненты.

### Планирование оцифровки и мобилизации данных

- Различные этапы планирования проекта по оцифровке данных и их адаптация к конкретному проекту.

### Получение данных

- Определение типов данных о биоразнообразии и способов их оцифровки с использованием лучших практик, существующие инструменты и методы.

### Управление данными

- Основные инструменты и концепции для проверки и верификации данных.

### Публикация данных

- Обеспечение свободного доступа к данным через Интернет (публикация данных) с использованием интегрированного инструмента GBIF Integrated Publishing Toolkit (IPT).

## В результате освоения курса вы сможете:

- Разработать стратегию мобилизации данных, адаптированную к конкретной институциональной структуре.
- Применить модель для построения стратегии мобилизации данных и использовать соответствующие протоколы оцифровки и сбора данных.
- Оценить стратегию мобилизации данных для выявления потенциальных пробелов, неэффективности и ловушек.
- Применять протокол оцифровки для получения цифровых данных о биоразнообразии из аналоговых источников.
- Использовать программные средства, предназначенные для облегчения оцифровки данных о биоразнообразии.
- Применять протоколы проверки и верификации данных о биоразнообразии для оценки и повышения их пригодности для использования.
- Использовать программные средства для оценки пригодности для использования набора данных о биоразнообразии.
- Использовать программные средства, предназначенные для проверки и верификации данных о биоразнообразии.
- Применять механизм публикации данных о биоразнообразии.
- Определять публикуемые типы данных и подтипы (если таковые имеются) наборов данных о биоразнообразии.
- Использовать GBIF IPT для публикации наборов данных о биоразнообразии с использованием соответствующих расширений.
- Помогать другими исследователям в оцифровке, управлении и публикации данных о биоразнообразии.
- Принесите свой собственный ноутбук на выездной курс для лучшего обучения.

## Преподаватели

### Local organizers

- Oleg Borodin, GBIF Беларуси, Национальная академия наук Беларуси, Минск, Беларусь
- Nina Voronova, Национальная академия наук Беларуси, Минск, Беларусь

### Учителя

- Dag Endresen - GBIF Норвегия, Университет Осло, Норвегия
- Piotr Tykarski - GBIF Польша, Университет Варшавы, Польша
- Dmitry Schigel - Секретариат GBIF, Копенгаген, Дания
- Maxim Shashkov - GBIF.ru, Институт физико-химических и биологических проблем почвоведения РАН, Пущино, Россия
- Natalya Ivanova - GBIF.ru, Институт математических проблем биологии РАН - филиал ИПМ им. М.В. Келдыша РАН, Пущино, Россия
- Laura Russell - Секретариат GBIF, Копенгаген, Дания (*внешняя поддержка*)

### Менторы

- Ana Sheiko Анна Шейко
- Hleb Kryshyn Греб Кришин

- Stanislaw Krasouski Станислав Красовский
- Yauheni Sysoliatin Евгений Сысолятин

## Программные инструменты

- Платформа онлайн-обучения: <http://elearning.gbif.es/courses/GBIFDMBIODATA2019/>
- IPT - Интегрированный инструмент для публикации данных: <https://ipt-biodata.gbif.no/belarus/>
- iNaturalist: <https://www.inaturalist.org/projects/biodata-belarus-2019>

## Язык

Русский и английский.

## Сертификаты о полученной квалификации

После успешной оценки и завершения курса участники будут иметь возможность получить официальный сертификат в виде значка GBIF Mozilla Open Badge.

## Предпосылки для участия в курсе

Чтобы успешно освоить материалы курса, участники должны обладать следующими знаниями и навыками:

- Базовые навыки использования компьютера и Интернета, в частности использования электронных таблиц, баз данных и инструментов для представления географических данных (например, карты Google, ГИС).
- Базовые знания о географии и информатике биоразнообразия: концепции географии и картографии, основные правила таксономии и номенклатуры, а также базовые знания о GBIF и других подобных инициативах, работающих в области информатики биоразнообразия.
- Хорошее владение английским языком. Занятия будут проводиться на английском языке. В курсе будут участвовать русскоязычные наставники, учебные материалы будут переведены на русский язык и доступны всем участникам.
- Готовность распространять знания, полученных в ходе курса, среди ваших партнеров и коллег в ваших проектах путем адаптации учебных материалов по мобилизации данных о биоразнообразии к конкретным условиям и языкам, сохраняя их учебную ценность.

## Мероприятия перед курсом: с 28 октября по 8 ноября 2019 года

В течение двух недель до начала очной части курса участники должны зарезервировать около 10 часов свободного времени для выполнения заданий онлайн. Все участникам будет предоставлен доступ к информации и учетные данные на платформе онлайн обучения [e-learning platform](#) и [учебной IPT](#). Материалы для самоподготовки доступны здесь: [модуль 01](#), [модуль 02](#).

### **Модуль 01: Введение в онлайн обучение**

Первый модуль подготовительных онлайн-мероприятий поможет участникам ознакомиться с онлайн-платформами, которые будут использоваться во время курса ([e-Learning platform](#)).

**Сессия 01a:** Введение в курс, обзор и практическая информация (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 01b:** Как использовать платформу e-learning (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 01c:** Введение в онлайн обучение (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 01d:** GBIF и другие инициативы (Подготовительное чтение) [[Слайды на русском](#)]

### **Модуль 02: Основные активности**

Этот модуль познакомит студентов с наукой информатикой биоразнообразия и стандартом Darwin Core.

**Сессия 02a:** Информатика биоразнообразия (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 02b:** Качество данных (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 02c:** Процесс оцифровки (Подготовительное чтение) [[Слайды на русском](#)]

**Сессия 02d:** Программные инструменты (Подготовительная работа) [[Слайды на русском](#)] ([Таблица ПО](#))

Пожалуйста, постарайтесь перед курсом установить программное обеспечение ([OpenRefine](#), [Java JRE](#)).

Мы, конечно же, поможем вам с установкой программного обеспечения во время семинара.

Пожалуйста, не забудьте **взять свой ноутбук** и убедитесь, что вы имеете **администраторские права** для установки необходимого программного обеспечения.

**Сессия 02e:** Список литературы для чтения перед курсом [[Слайды на русском](#)] ([Источники для чтения](#))

**Сессия 02f:** Опросник [[Russian slides](#)]

Пожалуйста, заполните [опросник](#) [[на русском](#)] до 9 ноября 2019.

## **Очная часть курса: 11–15 ноября 2019 года**

Материалы очной части курса доступны через платформу онлайн-обучения, а также [по ссылке](#). [Короткая программа](#) и программа (на [английском](#) и [русском](#)).

### **День 1 - Понедельник 11 ноября 2019**

#### **Модуль 03: ВВЕДЕНИЕ В ОЧНУЮ ЧАСТЬ КУРСА**

Этот модуль позволит участникам познакомиться друг с другом, с GBIF и с онлайн средствами, которые мы будем использовать во время курса, а также основными теоретическими концепциями, которые будут использоваться в разных модулях.

08:30 - Регистрация (30 минут)

09:00 - **Сессия 03a:** Добро пожаловать! Введение (60 минут) [[Слайды на русском](#)]

Эта сессия будет включать приветствие от организаторов курса, практическую информацию, которая нам понадобится во время курса; обзор деятельности перед курсом; и представление участников (презентации и введение).

#### **Модуль 04: ОСНОВЫ ИНФОРМАТИКИ БИОРАЗНООБРАЗИЯ**

В этом модуле представлены основные понятия и терминология, которые будут использоваться в учебных модулях курса.

10:00 - **Сессия 04a**: Основы: Терминология и стандарты (45+15 минут) [[Слайды на русском](#)]  
В ходе этой сессии мы обсудим ключевые концепции, которые мы будем использовать во всех модулях, в частности, об управлении цифровыми данными о биоразнообразии, а также концепции стандартов данных. Мы рассмотрим важность документирования и сосредоточимся на метаданных и подготовке данных к публикации (презентация и обсуждение).

11:00 - Кофе-брейк (30 минут)

11:30 - **Сессия 04b**: Основы: стандарт Darwin Core (30+15 минут) [[Слайды на русском](#)]  
Во время этой сессии участники познакомятся со стандартом Darwin Core и его компонентами, которые будут использоваться на протяжении оставшейся части курса (презентация и обсуждение).

12:15 - **Сессия 04c**: Основы: качество данных (30+15 минут) [[Слайды на русском](#)]  
Во время этой сессии участники познакомятся с общими принципами качества данных, которые будут использоваться на протяжении оставшейся части курса (презентация и обсуждение).

13:00 - Обед (60 минут)

#### **Модуль 05: ПЛАНИРОВАНИЕ**

Этот модуль знакомит участников с методами планирования проекта по мобилизации данных о биоразнообразии. Документ с описанием [ролей](#) и [этапов](#) (на английском и русском языках) доступен для карточной игры.

14:00 - **Сессия 05a**: Планирование: рабочий процесс и документирование (45 минут) [[Слайды на русском](#)]  
В этой сессии описываются основные этапы планирования успешного проекта по мобилизации данных и способы создания жизнеспособного рабочего процесса. Рассматриваемые вопросы: Какие задачи должны быть формально структурированы в начале? Что ожидается и что должно получиться? Кому нужно знать, что и когда ему делать? В этой сессии будет использована Ситуация 1 (презентация).

14:45 - **Сессия 05b**: Планирование: определение заинтересованных сторон и ролей (45 минут) [[Слайды на русском](#)]  
Практическая групповая сессия. Используя методы, обсуждавшиеся на предыдущем занятии, группам будет поручено создать идеализированный план проекта / рабочего процесса на основе Ситуации 1. Менторы могут, если это уместно, визуализировать собственный проект группы (упражнение).

15:30 - Кофе-брейк (30 минут)

16:00 - **Сессия 05c**: Планирование: кто, что и когда делает - задачи и этапы (30 минут)  
Практическая групповая сессия. Используя методы, обсужденные в предыдущей сессии, группам будет поручено создать план реальный план проекта / рабочего процесса. Затем они будут представлены всей группе на следующей сессии для обсуждения общих черт и различий (упражнение).

16:30 - **Сессия 05d**: Планирование: предлагаемое решение (60 минут) [[Слайды на русском](#)]  
Группы представляют аудитории результаты своих упражнений. Докладчик рассказывает о предлагаемом решении (обсуждение)

17:30 - Окончание дня

## День 2 - Вторник 12 ноября 2019

09:00 - [Сессия 03b](#): Зачем делиться данными? (60 минут) [[Слайды на русском](#)]

На этой сессии участники узнают, почему мы должны делиться данными (презентация и обсуждение).

### [Модуль 06](#): ПОЛУЧЕНИЕ ДАННЫХ О БИОРАЗНООБРАЗИИ

Этот модуль знакомит участников с типами данных о биоразнообразии и предоставляет лучшие практики для сбора данных.

10:00 - [Сессия 06a](#): Происхождение и типы данных о биоразнообразии (45+15 минут) [[Слайды на русском](#)]

Практическое занятие по определению с какими типами данных работают участники, и полезная информация для определения приоритетов сбора данных. Включает групповые обсуждения для ознакомления с различными видами данных (коллекции, наблюдения, сборы, таксономические списки, медиа), а также о том, как наилучшим образом делиться этой информацией (презентация и обсуждение).

11:00 - Кофе-брейк (30 минут)

11:30 - [Сессия 06b](#): Получение, обработка и качество данных (45 минут) [[Слайды на русском](#)]

Практическое занятие фокусируется на сборе данных с использованием ранее описанной Ситуации (презентация и обсуждение).

12:15 - [Сессия 06c](#): Получение, обработка и качество данных (45 минут) [Russian exercise]

Практическое занятие фокусируется на сборе данных с использованием ранее описанной Ситуации (упражнение).

13:00 - Обед (60 минут)

### [Модуль 07](#): ПРОВЕРКА, ВЕРИФИКАЦИЯ И СТАНДАРТИЗАЦИЯ ДАННЫХ

Этот модуль познакомит участников с основными инструментами и концепциями, используемыми для проверки и верификации данных, а также с тем, как данные могут быть стандартизированы для публикации в соответствии Darwin Core.

14:00 - [Сессия 07a](#): Принципы проверки и верификации данных (45 минут) [[Слайды на русском](#)]

Первая часть этого занятия познакомит участников с основными концепциями, связанными инструментами и лучшими практиками по проверке, верификации и стандартизации данных. Затем следует практическое упражнение по поиску и устранению технических ошибок и ошибок соответствия (презентация и упражнение).

14:45 - [Сессия 07b](#): Другие инструменты для проверки и верификации данных (45 минут) [[Слайды на русском](#)]

На этом занятии рассматриваются инструменты, используемые для проверки и верификации наборов данных в трех основных категориях ошибок: номенклатурные, формата и географические. Затем следует практическое упражнение (презентация и упражнение).

15:30 - Кофе-брейк (30 минут)

16:00 - [Сессия 07c](#): Проверка и верификация данных с использованием OpenRefine (30+60 минут) [[Слайды на русском](#)]

Презентация OpenRefine - простого инструмент для стандартизации и улучшения качества наборов данных. Затем следует практическое упражнение с использованием функций по умолчанию, существующих веб-сервисов и регулярных выражений (презентация и упражнение).

17:30 - Окончание дня

## День 3 - Среда 13 ноября 2019

Экскурсия: полевая экскурсия - [BioBlitz](#) с [iNaturalist](#)

Каждый учебный курс BioDATA включает биоблиц с введением в платформу [iNaturalist](https://www.inaturalist.org/projects/biodata-belarus-2019) для представления данных наблюдений за видами, <https://www.inaturalist.org/projects/biodata-belarus-2019>

Перед полевой поездкой и биоблицем вам следует проконсультироваться с сеансом [самообучения 6с](#).

09:00 - Посещение Беловежской пуши, прогулка в окрестностях места проведения курса. Всех студентов просят зарегистрировать собственную учетную запись пользователя на платформе iNaturalist и присоединиться к проекту bioblitz курса.

14:00 - Обед (60 минут)

15:00 - [Сессия 04d](#): Основные: документирование (30 минут) [[Слайды на русском](#)]

Во время этой сессии участники познакомятся с лучшими практиками документирования данных, которые будут использоваться на протяжении оставшейся части курса (презентации).

15:30 - Кофе-брейк (30 минут)

### [Модуль 08](#): ПУБЛИКАЦИЯ ДАННЫХ

Этот модуль посвящен процессу обеспечения свободного доступа к данным биоразнообразия в Интернете. Мы будем использовать интегрированный инструмент для публикации данных GBIF Integrated Publishing Toolkit (IPT).

16:00 - [Сессия 08d](#): Публикация данных со статьей о данных (data papers) (45 минут) [[Слайды на русском](#)]

Статьи о данных - это рецензируемые научные публикации, описывающие наборы данных, они важны для повышения пригодности для повторного использования данных другими исследователями. (Презентация)

16:45 - [Сессия 08e](#): FAIR открытые данные (45 минут) [[Слайды на русском](#)]

FAIR-данные, совместимые и доступные для поиска, могут использоваться повторно. Какова ценность FAIR-принципов данных для ваших наборов данных (презентация и обсуждение).

17:30 - Окончание дня

## День 4 - Четверг 14 ноября 2019

### [Модуль 08](#): ПУБЛИКАЦИЯ ДАННЫХ (через [IPT](#))

Этот модуль посвящен процессу обеспечения свободного доступа к данным о биоразнообразии через Интернет. Мы будем использовать интегрированный инструмент для публикации данных - GBIF IPT.

09:00 - [Сессия 08a](#): Публикация данных через IPT (90 минут) [[слайды на русском](#)]

Презентация по таким темам, как лицензии, метаданные, обязательные поля, размещение наборов данных различных учреждений на одной и той же установке IPT и т. д. Презентация и демонстрация, охватывающая основы публикации с использованием инструмента IPT (принципы, пользовательский интерфейс, рабочий процесс, метаданные, набор данных видимость и т. д.). Демонстрация и обсуждение, охватывающие функции [IPT](#), и публикация сложного набора данных на основе выборки, в котором будет сделан упор на использование расширений и отношения ядро / расширение (презентация и демонстрация).

11:00 - Кофе-брейк (30 минут)

11:30 - [Сессия 08b](#): Публикация данных через IPT - попробуйте сами (90 минут) [[Слайды на русском](#)]

Упражнения в группах, когда вы сами попытаетесь использовать [IPT](#) для публикации демонстрационных наборов данных или ваших собственных наборов данных в среде «песочницы», связанной с [демонстрационной версией портала GBIF](#) в Копенгагене (упражнение).

13:00 - Обед (60 минут)

14:00 - [Сессия 08c](#): Публикация данных через [IPT](#) - обсуждение (90 минут) [[Слайды на русском](#)]



Обсуждение упражнения <a href="#">IPT</a> и просмотр наборов данных, которые были опубликованы студенческими группами (презентация и обсуждение).
15:30 - Кофе-брейк (30 минут)
16:00 - Публикация данных через <a href="#">IPT</a> - продолжение (90 минут) [Слайды на русском] Упражнения в группах, когда вы сами пытались использовать <a href="#">IPT</a> для публикации демонстрационных наборов данных или ваших собственных наборов данных в среде «песочницы», связанной с <a href="#">демонстрационной версией портала GBIF в Копенгагене</a> (упражнение).
17:30 - Окончание дня

### День 5 - Пятница 15 ноября 2019

#### Упражнение с [Ситуацией 2](#)

Работа в группах

09:00 - Групповая работа - [Ситуация 2](#) (90 минут) [Слайды на русском]

Работа над групповым заданием (упражнение).

11:00 - Кофе-брейк (30 минут)

11:30 - Групповая работа: продолжение (90 минут)

Работа над групповым заданием (упражнение). Продолжение

13:00 - Обед (60 минут)

14:00 - Презентации, Обсуждение (90 минут)

Обсуждение результатов выполнения группового задания.

15:30 - Кофе-брейк (30 минут)

#### [Модуль 09](#): ЗАКЛЮЧЕНИЕ ОЧНОЙ ЧАСТИ КУРСА

Во время этого последнего модуля местного курса мы рассмотрим содержание и обсудим последующие действия. Участники будут иметь возможность представить свою оценку курса. Вручение сертификатов участника.

16:00 - [Session 09](#): Заключение очной части курса и подведение итогов (30 минут) [Слайды на русском]

Презентация, дискуссия и [оценочная форма](#) [на русском].

16:30 - Окончание курса



## Самостоятельная работа после очной части курса: С 25 ноября по 6 декабря 2019 года

Участники должны зарезервировать около 10 часов свободного времени для выполнения заданий в течение двух недель после очной части курса.

**Модуль 10:** Подготовка набора данных по [Ситуации 3](#) и его публикация через IPT

*Выполнение этой части дает возможность участникам продемонстрировать знания и навыки, полученные ими при освоении курса. Результаты выполнения этих упражнений станут основой для оценки участников по системе ECTS и сертификации GBIF.*

**Сессия 10а:** GBIF портал [[Слайды на русском](#)]

**Сессия 10б:** Использование данных GBIF [[Слайды на русском](#)]

**Сессия 10с:** Итоговая оценка и [Ситуация 3](#) [[Слайды на русском](#)]

## КОРОТКАЯ ПРОГРАММА

День 1 Понедельник, 11 ноябрь	День 2 Вторник, 12 ноябрь	День 3 Среда, 13 ноябрь	День 4 Четверг, 14 ноябрь	День 5 Пятница, 15 ноябрь
09:00: Презентация <b>03a.</b> Введение и участие в GBIF	09:00: Презентация / обсуждение <b>03b.</b> Основы: зачем делиться данными?	09:00: Полевая экскурсия / биоблиц Полевая экскурсия с <b>биоблицем</b> (и <b>iNaturalist</b> )	09:00: Презентация / демо <b>08a.</b> Публикация данных через <b>IPT</b>	09:00: Упражнение Упражнение с <b>Ситуацией 2</b>
10:00: Презентация <b>04a.</b> Основы: терминология и стандарты	09:45: Презентация / обсуждение <b>06a.</b> Происхождение и типы данных о биоразнообразии	Сеанс <b>самообучения 06с</b> рекомендуется перед поездкой <b>06с.</b> Сбор данных с <b>iNaturalist</b>		
11:00 Кофе-брейк	11:00 Кофе-брейк		11:00 Кофе-брейк	11:00 Кофе-брейк
11:30: Презентация <b>04b.</b> Основы: Darwin Core	11:30: Презентация / упражнение <b>06b.</b> Получение, обработка и качество данных	Полевая экскурсия	11:30: Упражнение <b>08b.</b> Публикация данных через <b>IPT</b> - попробуйте сами	11:30: Групповые презентации Упражнение с <b>ситуацией 2</b> продолжение
12:15: Презентация <b>04c</b> Основы: качество данных				
13:00 Обед	13:00 Обед	14:00 Обед	13:00 Обед	13:00 Обед
14:00: Презентация <b>05a.</b> Планирование: рабочий процесс и документирование	14:00: Презентация, упражнение <b>07a.</b> Базовые концепции проверки и верификации данных	15:00: Презентация <b>04d.</b> Основы: Документирование	14:00: Презентация / обсуждение <b>08с.</b> Публикация данных через <b>IPT</b>	14:00: Обсуждение Упражнение с <b>Ситуацией 2</b> (продолжение)
15:00: Упражнение <b>05b.</b> Планирование: заинтересованные стороны и роли	14:45: Презентация, упражнение <b>07b.</b> Проверка и верификация данных с использованием других инструментов			
15:30 Кофе-брейк	15:30 Кофе-брейк	15:30 Кофе-брейк	15:30 Кофе-брейк	15:00 Кофе-брейк
16:00: Упражнение <b>05с.</b> Планирование: задачи и стадии	16:00: Презентация / упражнение <b>07с.</b> Проверка и верификация данных в OpenRefine	16:00: Презентация / обсуждение <b>08d.</b> Статьи о данных	16:00: Упражнение Публикация данных через <b>IPT</b> - (продолжение)	15:30: Заключение <b>09.</b> Заключение и подведение итогов
16:30: Обсуждение <b>05d.</b> Планирование: Предлагаемые решения		16:45: Презентация / обсуждение <b>08e.</b> FAIR данные		Final follow-up assignment Follow-up task <b>Use Case 3</b>
17:30 Окончание дня	17:30 Окончание дня	17:30 Окончание дня	17:30 Окончание дня	16:00 Окончание дня

**Место встречи:** [Беловежская пушча, Каменюки](#), Беларусь 11-15 ноября 2019 г. [Домашняя страница курса здесь](#).

Подготовительные материалы: [модуль 01](#), [модуль 02](#). IPT <https://ipt-biodata.gbif.no/belarus/>

Короткая программа обновлена 2019-11-07. [EN](#) & [RU](#)