

**План индивидуального проекта
по курсу “Цифровые технологии в исследовании биоразнообразия”**

Проект нужно оформить в виде текстового файла и прислать в телеграм-группу
до 30 ноября 2022 г.

**Название: Данные о [название целевого вида] в международных
репозиториях о биоразнообразии и их характеристика**

Обязательная часть

Теоретическая (объем не более 1 страницы)

1. Обоснование выбора целевого вида и территории (1 абзац, почему выбран именно этот вид и территория).
2. Краткое описание биологии и экологии вида (1 абзац, не забудьте указать ссылки на использованные источники). Например, образ жизни, предпочитаемые биотопы, способы размножения и расселения, положение в трофической пирамиде, суточный и сезонный циклы активности. Есть ли у выбранного вида статус IUCN, является ли он инвазивным?
3. Описание ареала вида в природе (1 абзац, можно вставить карту, если она доступна). Насколько широко распространен этот вид (является ли эндемиком какой-то территории), расширяется ли его ареал или, наоборот, сокращается.

Практическая (объем не более 1 страницы)

1. Из репозитория GBIF выгрузите все записи о находках целевого вида для выбранной территории. Укажите DOI сформированного набора данных, приведите ссылку на него в списке литературы.
2. Опишите полученные данные по следующей схеме:
 - Из каких наборов данных происходят? Можно составить таблицу с названием набора данных и числом записей.
 - Какой процент записей имеет географическую привязку? Есть ли оценка точности привязки, в каком диапазоне она находится?
 - В какие годы собраны полученные данные?
 - Насколько данные GBIF соответствуют литературным сведениям об ареале целевого вида?

Творческая часть

Обязательно нужно выполнить одно любое задание из этой части, остальные - по желанию.

1. **Работа с QGIS.** Постройте карту распространения целевого вида в QGIS. Опишите насколько полно найденные находки отражают ареал вида, насколько равномерно или наоборот агрегировано они распределены, есть ли точки далеко за пределами ареала (есть ли уверенность в этих точках или это ошибки привязки).
2. **Оценка качества данных при помощи электронных таблиц.** При помощи электронных таблиц найдите и исправьте ошибки в выгруженных данных. В качестве подсказки можно использовать сведения об автоматически найденных несоответствиях, обнаруженных при формировании поискового запроса в GBIF. В тексте опишите, какие ошибки вы нашли, в каких строках (номера строк) сделаны исправления и какие функции Open Refine использовали для исправления ошибок. Приложите 2 CSV файла - до исправлений и после.
3. **Оценка качества данных при помощи Open Refine.** При помощи программы OpenRefine найдите и исправьте ошибки в выгруженных данных. В качестве подсказки можно использовать сведения об автоматически найденных несоответствиях, обнаруженных при формировании поискового запроса в GBIF. В тексте опишите, какие ошибки вы нашли, в каких строках (номера строк) сделаны исправления и какие функции Open Refine использовали для исправления ошибок. Приложите 2 CSV файла - до исправлений и после.
4. **Работа с R.** Постройте график, описывающий загруженные из GBIF данные. Например, это может быть столбчатая диаграмма (barplot) с число записей по годам (временным периодам). Помимо рисунка вставьте в текст работы программный код с комментариями и приложите файл с исходными данными.
5. **Построение прогнозного ареала целевого вида.** Соответствует ли он вашим представлениям о распространении данного вида? Есть ли области где вероятность его встретить предсказана явно ошибочно и наоборот, "пропущенные" территории, где этот вид по вашему мнению должен быть точно. Какие еще переменные среды следовало бы учесть для построения более адекватной модели. Вставьте в текст работы программный код с комментариями.

Список литературы

Не забудьте вставить в текст работы ссылки на все использованные источники и привести их в Списке литературы.