

Оцифровка и мобилизация данных находок грибов из литературы по территории Западной Сибири: обзор и результаты 5-летней работы

Научный руководитель – Филиппова Нина Владимировна

Тюнина Лина Андреевна

Студент (бакалавр)

Югорский государственный университет, Ханты-Мансийск, Россия

E-mail: linatyunina2309@icloud.com

Интеграция данных о находках видов в формат открытых баз данных имеет огромное значение для научных и практических задач изучения и охраны биоразнообразия. Все операции по оцифровке и публикации данных на открытых порталах именуются «мобилизацией» данных, что делает их доступными и мобильными для использования. Пять лет назад нами начата инициатива интеграции данных о находках грибов и грибоподобных организмов из опубликованных ранее источников на территории Западной Сибири. В настоящее время мобилизовано около половины доступных в литературе данных. В тезисах представлены результаты развития проекта на начало 2023 года.

В работе по мобилизации данных о находках грибов из литературных источников был принят протокол работы, включающий несколько стадий:

1. Создание библиографической базы данных и библиотеки публикаций, предназначенных для оцифровки и мобилизации данных.
 - a. Список работ составляется на основе библиографий авторов, работающих на территории, их предшественников, основных тематических журналов и пр.
 - b. База данных библиографии ведется в библиографическом менеджере Zotero и доступна в облачном сервисе для авторизованных пользователей.
 - c. Полные тексты источников сканируются и сохраняются в базе данных.
 - d. Пополнение базы данных осуществляется ее кураторами.
2. Ведение протокола оцифровки источников.
 - a. Для отслеживания прогресса оцифровки источников создана таблица-лог оцифровки.
 - b. В таблице содержится информация о библиографическом источнике, качестве представленных в нем данных, процессе оцифровки и качестве оцифрованных данных.
 - c. Заполнение протокола оцифровки осуществляется его куратором.
3. Данные вносятся в подготовленные шаблоны (в формате Darwin Core) в гугл-таблицах (google sheets) или Microsoft Excel.
 - a. Шаблон представляет собой простую таблицу из 24 столбцов, включающих основную информацию о находке и ее источнике.
4. Подготовленные шаблоны по одному или нескольким источникам проходят проверку качества куратором базы данных и интегрируются в общую базу данных.
5. База данных регулярно (1 раз в 2-3 месяца) обновляется в GBIF в ручном режиме (импортом новой таблицы csv с настольного компьютера).

Будущие планируемые усовершенствования протокола включают:

1. Импорт базы данных литературных находок в ИС менеджмента коллекций "Specify".
2. Автоматический экспорт/обновление данных из ИС в GBIF через MySQL запрос к базе.

В результате проведенной работы по оцифровке и мобилизации находок грибов на территории Западной Сибири в открытом доступе представлено около 30 тысяч точек находок. Данные могут быть скачаны одним массивом по ссылке (<https://doi.org/10.15468/dl.ed4es9>) или воспользовавшись фильтром по таксонам, географии или другим параметрам (рис.). Данные объединены по географическому принципу в виде двух наборов данных GBIF - по северу и югу Западной Сибири [1, 2, 3, 4].

Источники и литература

- 1) Рабочая группа оцифровки литературных данных. Сибирское Микологическое Общество. Дата доступа: 17.02.2023. URL: <https://sibmyco.org/literaturedatabase/>
- 2) Fungal literature records database of the Northern West Siberia (Russia). Yugra State University Biological Collection (YSU BC). Accessed via GBIF.org on 2023-02-17. URL: <https://doi.org/10.15468/hfje3l>
- 3) Fungal literature records database of West Siberia. Yugra State University Biological Collection (YSU BC). Дата доступа: 17.02.2023. URL: <https://fungariumysu.org/funws/>
- 4) The Fungal Literature-based Occurrence Database for the Southern West Siberia (Russia). Yugra State University Biological Collection (YSU BC). Accessed via GBIF.org on 2023-02-17. URL: <https://doi.org/10.15468/eqx72v>

Иллюстрации

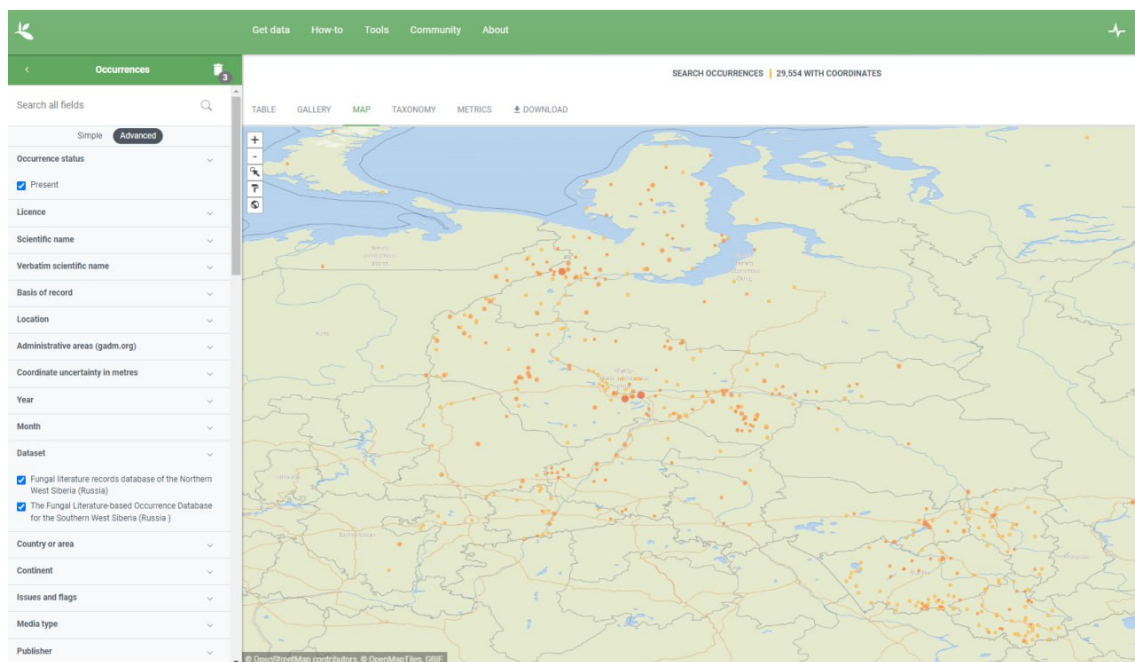


Рис. Скриншот карты оцифрованных находок грибов на территории Западной Сибири, представленных в GBIF в виде двух наборов данных (в правом меню помечены галочкой).