

Современное состояние и перспективы развития цифровой коллекции «Виртуальный Гербарий ALTB»

Current status and development prospects of the digital collection «Virtual Herbarium ALTB»

Ваганов А. В., Медведева К. Е.

Vaganov A. V., Medvedeva K. E.

Алтайский государственный университет, г. Барнаул, Россия, E-mail: vaganov_vav@mail.ru

Altai State University, Barnaul, Russia

Реферат. В работе представлено современное состояние дел и перспективы развития цифровой коллекции «Виртуальный Гербарий ALTB», созданной в 2010 году. Произведен анализ эффективности, решаемых при помощи цифровой коллекции задач проекта «Флора Алтай». Оценены наиболее востребованные направления развития при взаимодействии с Глобальной информационной системой по биоразнообразию (GBIF). Приведены доводы о необходимости интеграционного взаимодействия трех субъектов процесса развития цифровой коллекции «Виртуальный Гербарий ALTB» – проект «Флора Алтай», GBIF и открытые спецификации данных (Darwin Core и GeoJSON).

Ключевые слова. Алтайская горная страна, база данных, большие данные, Гербарий, ГИС, типовой материал, фиторазнообразие, флора, ALTB, Darwin Core, GBIF, GeoJSON.

Summary. The paper presents the current state of affairs and prospects for the development of the digital collection «Virtual Herbarium ALTB», created in 2010. The effectiveness of solving tasks using the digital collection within the project «Flora Altaica» has been analyzed. The most demanded directions of development in cooperation with the Global Biodiversity Information Facility (GBIF) are estimated. Arguments are given about the necessity of integration interaction between the three actors in the development of the digital collection «ALTB Virtual Herbarium» – the project «Flora Altaica», GBIF, and open data specifications (Darwin Core and GeoJSON).

Kew words. Altai mountain country, ALTB, big data, Darwin Core, database, flora, GBIF, GeoJSON, GIS, Herbarium, phytodiversity, type material.

Гербарий Южно-Сибирского ботанического сада Алтайского государственного университета с 1996 г. включен в международный Союз гербариев мира (Index Herbariorum) под акронимом ALTB. На сегодняшний день это один из самых динамично развивающихся гербариев в России. По количеству числа гербарных образцов он занимает седьмое место в России среди университетских коллекций и коллекций институтов Академии наук (более 450 000 гербарных листов). Среди всех гербарных коллекций мира фонд Гербария ALTB является самым крупным по трансграничной территории Алтайской горной страны (АГС).

Оцифровка естественнонаучных коллекций является мощным инструментом для систематизации коллекционных фондов и повышения эффективности управления ими через быстрый доступ к образцам и связанной с ними информации. Наличие изображений коллекционных образцов и полных сведений о них в Интернете повышает востребованность коллекций, чаще вовлекая тот или иной образец в научный оборот (Серегин, 2018). Доступ к данным позволяет избежать длительной переписки с кураторами коллекций и экономит время и ресурсы исследователя (Шашков и др., 2018).

В России процесс перевода в цифровой формат национальных биологических коллекций находится на начальном этапе (Филиппова и др., 2017). В нашей стране зарегистрировано 116 гербарных коллекций с общим фондом около 16 млн образцов, из них по данным на сентябрь 2019 г. (Ковтонюк, 2019).

В 2010 г. на основе гербарной коллекции АЛТВ был разработан «Виртуальный Гербарий АЛТВ», который позволяет, в первую очередь, сохранять реальный гербарный образец, что особенно важно в отношении типового материала (Vaganov et al., 2018). Такая необходимость в создании виртуального Гербария была продиктована мировыми трендами по цифровизации коллекций и определена задачами ботанических садов. В частности, в задачи Южно-Сибирского ботанического сада входят: подготовка и издание многотомной сводки «Флора Алтай»; исследование флоры и растительного покрова Алтайской горной страны; монографическое изучение отдельных групп растений; проведение лабораторных занятий, летних практик, специальных курсов; разработка мер по охране редких и исчезающих видов растений (Ваганов, 2011).

Подробное описание структуры базы данных «Виртуальный Гербарий АЛТВ», технологии накопления и обработки текстовой и графической информации описаны автором настоящей работы ранее (Ваганов, 2011). Произведя анализ ресурса за почти 10-летний период, можно отметить, что Гербарий востребован не только на национальном уровне, но и за его пределами. При этом цифровая коллекция востребована наиболее широко по следующим направлениям: удаленный доступ изображений гербарных листов, уточнение местообитаний таксонов, а также для учебно-исследовательского материала обучающихся, подготовки курсовых и дипломных работ. Отдельно отметим, что стало обязательным наличие в цифровой коллекции гербарных листов, с которых были взяты образцы для молекулярно-генетического анализа, либо споры или пыльца для применения методов сканирующей электронной микроскопии. Инвентаризационный номер при этом указывается в самой научной статье, полученной по итогам исследования, а также в генбанках.

Если взглянуть на тему развития цифровых Гербариев шире, в более глобальных масштабах, то в настоящее время большинство мировых коллекций в цифровом формате переведены из локальных баз данных в Сеть Интернет и по последним данным более 200 их них индексируются в GBIF.org (Global Biodiversity Information Facility – Глобальная информационная система по биоразнообразию). GBIF является основой для облегчения оцифровки данных о биоразнообразии и для обеспечения совместимости пока еще неизвестного числа баз данных о биоразнообразии, которые распространены по всему миру (Edwards et al., 2000). В Алтайском государственном университете процесс интеграции данных в GBIF начался сравнительно недавно. Основным набором данных для GBIF от АЛТГУ является Гербарий АЛТВ. В рамках партнерского соглашения АЛТГУ с GBIF (www.gbif.org/publisher/943a5811-d56e-4c37-853d-bd64957d3833) в период с 2018 по 2019 гг. было предоставлено 17350 этикеточных записей, подкрепленных цифровым изображением гербарного листа (тип данных – «OccurrenceData», сведения об отдельных находках) (Vaganov et al., 2019).

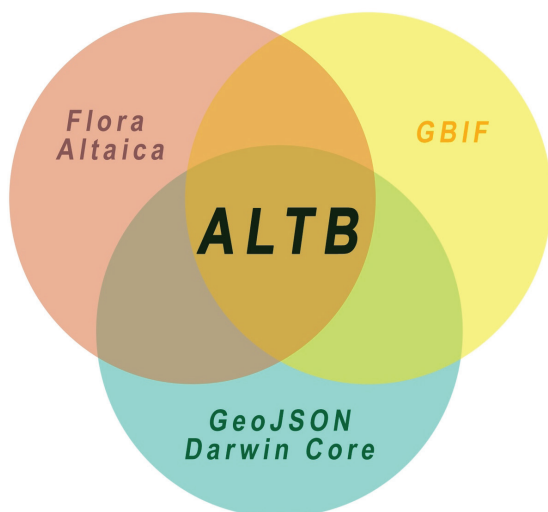


Рис. Схема взаимодействия трех субъектов развития цифрового Гербария АЛТВ.

Процесс работы с GBIF АЛТГУ, как издателя данных о биоразнообразии, позволил расширить возможности цифровой коллекции АЛТВ. Проект «Флора Алтай», GBIF и спецификации данных (Darwin Core, GeoJSON) оказались тесно связаны между собой, имеют близкие цели и задачи. На данном уровне работы по обмену, хранению и обработке данных о фито-разнообразии стали более эффективными, что в итоге выражается путем открытой интеграции научных депозитариев в GBIF (рис.).

Спецификация Darwin Core, разработанная международной группой «International Working Group on Taxonomic Databases For Plant Sciences (TDWG)», предназначена для хранения данных о биоразнообразии в мире, способствует объединению информации, происходящей из разных источников (Wieczorek et al., 2012). Ис-

пользование единого стандарта представления данных и свободный доступ к ним позволяет использовать эту информацию для анализа в составе объединенных массивов (Шашков и др., 2018). Спецификация Darwin Core стандартизовала обмен данными о фиторазнообразии, а спецификация географических структур GeoJSON позволила произвести открытое районирование территории АГС для всех заинтересованных пользователей. Данные спецификации позволили заложить новые методологические стандарты в работе и привнести методы информатики биоразнообразия в выпускные квалификационные работы и диссертационные исследования.

Для объединения ботаников, изучающих флору Алтайской горной страны, в 2018 г. открыт проект «Флора Алтай» (<http://altaiflora.asu.ru>). Разделы проекта содержат, в том числе, оригинальные макеты печатной версии «Флора Алтай», ключи таксонов и карту АГС. Карта имеет 19 ботанико-географических районов АГС (полигонов) в спецификации GeoJSON (Ваганов и др., 2019). Использование полигонов GeoJSON, а также шейп-фалов, полученных на их основе, позволяют подготавливать карты распространения в ГИС-программах по территории АГС, уточнять современный ареал местообитаний таксонов, а также автоматизировано готовить точные указания местонахождений для 19-ти ботанико-географических районов в конспектах. Полигоны АГС доступны для скачивания всем заинтересованным пользователям в разделе «Карта АГС» проекта.

Сеть GBIF объединяет все источники посредством использования спецификации Darwin Core. Для инвентаризации коллекции и отображения этикеточной информации ALTB мы использовали 20 терминов из Darwin Core: occurrenceID, references, basisOfRecord, country, countryCode, family, genus, specificEpithet, verbatimElevation, recordedBy, acceptedNameUsage, catalogNumber, verbatimLocality, verbatimCoordinates, decimalLatitude, decimalLongitude, verbatimEventDate, eventDate, identifiedBy, typeStatus. Гипертаблица этикеточных данных регулярно пополняются для поддержания Виртуального Гербария, а данные интегрируются в GBIF посредством узла IPT (The Integrated Publishing Toolkit).

Существенной проблемой при инвентаризации коллекций, особенно для старых Гербариев, является отсутствие указания координат в этикетках, что существенно понижает ценность коллекции, в первую очередь, для составления ареалов таксонов. Следовательно, после этапа оцифровки коллекции и внесения этикеточных данных следует кропотливая работа по дополнительным геопривязкам и работа с картами. Выполняя дополнительные геопривязки, как правило, нам удается повышать процент данных с координатами с ~50 % до ~70 % от общего оцифрованного объема. Финальными действиями по поддержанию виртуальной коллекции ALTB являются конечная адаптация данных этикеток под спецификацию Darwin Core и последующая интеграция через IPT в GBIF.

Таким образом, всю современную работу по развитию цифровой коллекции «Виртуальный Гербарий ALTB» невозможно рассматривать без участия трех важных субъектов – проект «Флора Алтай», GBIF и спецификации данных Darwin Core и GeoJSON, каждый из которых вносит свой существенный вклад в ее развитие. Благодаря им этапы работы с гербарным фондом перешли на новый более глобальный, качественный и автоматизированный уровень.

Благодарности. Исследование выполнено при финансовой поддержке РФФИ и Алтайского края в рамках научного проекта № 19-44-220004_p_a.

ЛИТЕРАТУРА

Ваганов А. В. Виртуальный гербарий Южно-Сибирского ботанического сада (ALTB) // Проблемы ботаники Южной Сибири и Монголии // Материалы X междунар. науч.-практ. конф. (Барнаул, 24–27 октября 2011 г.) – Барнаул, 2011. – С. 27–28.

Ваганов А. В., Шмаков А. И., Гудкова П. Д. Глобальные данные о фиторазнообразии Алтайской горной страны, представленные в мировых научных депозитариях // Проблемы ботаники Южной Сибири и Монголии: сб. науч. ст. по материалам XVIII междунар. науч.-практ. конф. (20–23 мая 2019 г., Барнаул). – Барнаул: Изд-во АлтГУ, 2019. – С. 222–227. DOI: <https://doi.org/10.14258/pbssm.2019045>

Ковтонюк Н. К. Проблемы и перспективы цифровых гербарных коллекций России // Тезисы докл. Всерос. науч. конф. с междунар. участием «Инновации и традиции в современной ботанике», посвящ. 150-летию со дня

рождения академика В. Л. Комарова (48-е Комаровские чтения) (Санкт-Петербург, 21–25 октября 2019 г.). – СПб.: Ботанический институт им. В. Л. Комарова РАН, 2019. – С. 50.

Серегин А. П. Гербарий МГУ им. М. В. Ломоносова сегодня // Зоологические исследования, 2018. – № 20. – С. 137–138.

Шашков М. П., Иванова Н. В., Филиппова Н. В., Щигель Д. С. Возможности, решения и инструменты GBIF для оцифровки и развития естественнонаучных коллекций // Зоологические исследования, 2018. – № 20. – С. 169–174.

Филиппова Н. В., Филиппов И. В., Щигель Д. С., Иванова Н. В., Шашков М. П. Информатика биоразнообразия: мировые тенденции, состояние дел в России и развитие направления в Ханты-Мансийском автономном округе // Динамика окружающей среды и глобальные изменения климата, 2017. – 8 (2). – С. 45–56.

Vaganov A. V., Medvedeva K., Sinitsyna T. Complex of open access databases of Altai State University for the purposes of the Altai-Sayan ecoregion research // Abstracts of the conference «Information Technologies in the Research of Biodiversity» (BIT–2018). – Irkutsk, 2018. – Pp. 39–41.

Vaganov A., Shmakov A., Zaikov V., Zholnerova E., Shalimov A., Belkin D., Batkin A., Kasatkin D., Kosachev P., Antonyuk E., Medvedeva K., Usik N. (2019). Virtual Herbarium ALTB (South-Siberian Botanical Garden). Version 1.30. Altai State University. Occurrence dataset <https://doi.org/10.15468/y6xmme> accessed via GBIF.org on 2020-02-07.

Edwards J. L., Lane M. A., Nielsen E. S. Interoperability of Biodiversity Databases: Biodiversity Information on Every Desktop // Science, 2000. – Vol. 289. – Pp. 2312–2314. DOI: 10.1126/science.289.5488.2312

Wieczorek J., Bloom D., Guralnick R., Blum S., Döring M., Giovanni R., Robertson T., Vieglais D. Darwin Core: An Evolving Community-Developed Biodiversity Data Standard // PLoS ONE, 2012. – Vol. 7 (1). – P. e29715.